Multimedia Systems (2014) 20:253–281
DOI 10.1007/s00530-013-0318-0

REGULAR PAPER

Spatially aggregating spectral descriptors for nonrigid 3D shape retrieval: a comparative survey

Chunyuan Li · A. Ben Hamza

Received: 25 October 2012 / Accepted: 17 March 2013 / Published online: 4 April 2013

Abstract This paper presents a comprehensive review and analysis of recent spectral shape descriptors for nonrigid 3D shape retrieval. More specifically, we compare the latest spectral descriptors based on the Laplace–Beltrami (LB) operator, including ShapeDNA, heat kernel signature, scale invariant heat kernel signature, heat mean signature, wave kernel signature, and global point signature. We also include the eigenvalue descriptor (EVD), which is a geodesic distance-based shape signature. The global descriptors ShapeDNA and EVD are compared via the chi-squared distance, while all local descriptors are compared using the codebook model. Moreover, we investigate the ambiguity modeling of codebook for the densely distributed low-level shape descriptors. Inspired by the ability of spatial cues to improve discrimination between shapes, we also propose to adopt the isocontours of the second eigenfunction of the LB operator to perform surface partition, which can significantly ameliorate the retrieval performance of the time-scaled local descriptors. In addition, we introduce an intrinsic spatial pyramid matching approach in a bid to further enhance the retrieval accuracy. Extensive experiments are carried out on two 3D shape benchmarks to assess the performance of the spectral descriptors. Our proposed approach is shown to provide the best performance.

Keywords Shape retrieval · Spectral geometry · Intrinsic partition · Aggregate local descriptors

Communicated by P. Pala.

C. Li · A. Ben Hamza (✉)
Concordia Institute for Information Systems Engineering, Concordia University, Montréal, QC, Canada
e-mail: hamza@ciise.concordia.ca

1 Introduction

Recent advances in 3D imaging and processing, graphics hardware and networks have led to a whopping increase in geometry models available freely or commercially on the Web. As a result, the task of efficiently measuring the 3D object similarity to find and retrieve relevant objects for a given query and categorize an object into one of a set of classes has become of paramount importance in a wide range applications. The main challenge in 3D object retrieval algorithms is to compute an invariant shape descriptor that captures well the geometric and topological properties of a shape [1–5].

Content-based shape retrieval based on the comparison of shape properties is complicated by the fact that many 3D objects manifest rich variability, and invariance to different classes of transformations and shape variations is often required. One of the most challenging settings addressed is the case of nonrigid or deformable shapes, in which the class of transformations may be very wide due to the capability of such shapes to bend and assume different forms. Recently, various methods have been proposed to tackle nonrigid 3D shape recognition problem, particularly with the deformation invariant representation. These methods can be mainly categorized into two main classes: skeleton-based [6–10] and surface-based [11–15]. The former approaches usually capture the global topological structure of the shape, and a dissimilarity is often determined as the cost function to match two or more shapes. The latter methods, on the other hand, often represent a shape as a frequency histogram of deformation invariant local distances or vertex signatures. In this paper, we focus mainly on the second paradigm with local vertex descriptors. Research efforts on spectral shape analysis have recently resulted in numerous spectral descriptors [11–18],

which are predominately based on the Laplace–Beltrami (LB) operator [19, 20]. However, to date, no comprehensive comparison has been conducted in the literature, which often results in intractable situation when choosing appropriate descriptors for certain applications.

In the field of image processing, an analogous problem is image retrieval which refers to finding images depicting similar scenes or objects. Like 3D shapes, images may manifest significant variability. The computer vision and pattern recognition communities have recently witnessed a wide adoption of feature-based methods in object recognition and image retrieval applications. One popular family of feature-based techniques is the scale-invariant feature transform (SIFT) [21], which has shown a good performance in various scenarios [22]. Feature-based methods also allow to represent images as collections of "visual words" and treat them using text search approaches, such as the codebook model paradigm.

More recently, the authors in [23] explored analogous codebook model approaches applied to the problem of nonrigid 3D shape retrieval. They use the heat kernel signature (HKS) [11] and the scale-invariant heat kernel signature (SIHKS) [12] as "geometric words", and constructed shape descriptors by means of soft-assignment of visual words to the densely distributed vertex signatures. Low-level features may have a considerable effect on the recognition performance. Following the work in [23], we compare the recent spectral descriptors in the framework of the codebook model, and thus comprehensively analyze and recommend the descriptor which plays the same role as SIFT in the image domain. One inherent component of the codebook model is the assignment of discrete codewords to continuous low-level features. Despite the clear mismatch of this hard assignment with the nature of continuous features, the approach has been applied successfully to images. We explicitly model the codeword assignment ambiguity for the densely described 3D shape, which also provides an understanding of the different spectral descriptors.

For shape retrieval tasks, the codebook models, which represent a shape as an orderless collection of local features, have demonstrated impressive levels of performance [23]. However, because these models disregard all information about the spatial layout of the features, they have a limited descriptive ability. Several spatial extensions in the image domain have been proposed recently, the most widely used one is spatial pyramid matching [24]. Unfortunately, overcoming the spatial limitations in 3D shape analysis to build effective structural object descriptions is quite challenging, especially when the recognition system must be designed to work in the presence of large deformation changes. The direct approaches can use the existing consistent shape segmentation methods and geometric correspondence, but they achieve a good performance at a relatively high computational cost in addition to the fact the number of subregions cannot be fixed among different shape classes.

Inspired by the fact that the Reeb graph extracted from the second eigenfunction of the LB operator is pose independent and captures the global profile of surface geometry [25], we propose to adopt the level sets of this eigenfunction to intrinsically partition the surface. Since the construction of the second eigenfunction is an inseparable step in calculating the spectral descriptors, the proposed partition method is a natural ingredient of the current framework. Extensive experimental results show that the intrinsic partition significantly improves the retrieval accuracy of all the time-scaled spectral descriptors with varying codebook models. Moreover, the intrinsic spatial pyramid matching on surfaces is shown to be robust and yields the best results. In addition, the intrinsic spatial partition framework offers further insight into the success of these recently proposed spectral shape descriptors.

1.1 Contributions

The contributions of this paper may be summarized as follows:

1. We present a comprehensive survey and analysis of recent spectral descriptors for nonrigid 3D shape retrieval.
2. We investigate the ambiguity modeling of codebook for the densely distributed low-level shape descriptors.
3. We introduce the intrinsic spatial partition, which yields a significant retrieval accuracy improvement.

The rest of this paper is organized as follows. Section 2 provides a brief overview of some previous works pertinent to shape analysis and the codebook model. The deformation invariant shape representation is presented in Sect. 3, which starts by defining the LB operator on Riemannian manifolds, followed by its discretization and eigenanalysis. Section 4 briefly reviews the graph embedding and spectral shape descriptors. In Sect. 5, we describe the codebook model with various ambiguity methods. In Sect. 6, we propose the intrinsic spatial partition. Experimental results on two standard datasets are presented in Sect. 7. Finally, we conclude and point out future work directions in Sect. 8.

2 Previous work

Since the introduction of SIFT and the codebook model, image classification has witnessed a rapid and fruitful development in recent years. By contrast, the vast majority of 3D shape recognition methods are ad hoc, and the performance is usually limited due largely to two main

reasons: (1) no excellent descriptor like SIFT is available for 3D shapes; (2) an appropriate codebook model and its intrinsic spatial extension are challenging to find due to isometric shape transformation. In this section, we discuss some previous works and current developments pertinent to shape analysis and the codebook model.

2.1 Shape analysis

In recent years, considerable research efforts on shape analysis has been conducted in an bid to design a better shape descriptor aimed at finding the most relevant shapes. In the literature, there are several survey works [1–4] that have keen interest in systematic shape retrieval and the taxonomy. In the sequel, we present some developments on 3D shape analysis from early general shape description to recent spectral shape analysis.

2.1.1 Shape description

A 3D shape is usually represented as a volume or surface/mesh. Other effective representation methods are based on medial [8] or multiple views [26]. Over the past decade, there has been a flurry of research activity on surface-based shape recognition due largely to two key reasons: first, surface-based 3D models are more popular because of their highly effective representation ability and less memory storage. Second, humans are taught to differentiate between shapes mainly by surface features, and in many shape applications only the surface is of interest. Therefore, in this paper, we focus on surface-based shape recognition.

Early research works on 3D shape description have been centered primarily on invariance under global Euclidean transformations (i.e., rigid transformations). These works include the shape context [27, 28], shape distributions [29], and spherical harmonics [30]. Recently, significant efforts have been invested in exploring the invariance properties of shapes to nonrigid deformations. An intuitive approach is to replace the Euclidean distance with the geodesic one. The primary motivation is that unlike the Euclidean distance, which is basically a straight line between two points in 3D space, the geodesic distance captures the global nonlinear structure and the intrinsic geometry of the data. For example, Elad and Kimmel [31] computed a bending invariant signature of a surface by applying the multidimensional scaling procedure to the geodesic distance matrix. In [32], an information-theoretic framework using the geodesic shape distributions was proposed. Also, Jain et al. [33] constructed a shape descriptor for correspondence and retrieval [5] in the spectral domain of the geodesic distance matrix. The main drawback of the geodesic distance is that it suffers from strong sensitivity to topological noise, which might heavily damage the shape invariants.

2.1.2 Spectral shape analysis

The recently emerging field of diffusion geometry provides a generic framework for many methods in the analysis of geometric shapes [34]. It formulates the heat diffusion processes on manifolds. Spectral shape analysis is a methodology that relies on the eigensystem (eigenvalues and/or eigenfunctions) of the LB operator to compare and analyze geometric shapes. Levy [35] showed that the eigenfunctions can be well adapted to the geometry and the topology of an object. Coifman and Lafon [34] constructed diffusion distances as the L_2-norm difference of energy distribution between two points initialized with unit impulse functions after a given time. Through the statistic of the distribution, the spectral distances can also be used for nonrigid shape recognition [36]. Other similar spectral distances include the commute time distance [37] and the biharmonic distance [38]. Since the eigensystem of the LB operator is isometric invariant, it is well suited for the analysis and retrieval of nonrigid shapes, and it is more robust than the geodesic distance. By integrating the local distribution of features, the Intrinsic Shape Context was proposed in [28] as a natural extension of the 2D Shape Context to 3D nonrigid surfaces, and it was shown to outperform individual vertex descriptors in 3D shape matching.

A recent survey [39] on spectral mesh processing comprehensively reports the spectral methods derived from certain appropriately defined mesh operators and their applications. In this paper, however, we theoretically and experimentally review and compare spectral signatures based on the LB operator, including the HKS [11], SIHKS [12], heat mean signature (HMS) [13], wave kernel signature (WKS) [14], and global point signature (GPS) [15]. The details are provided in Sect. 4.

2.2 Codebook model

The past decade has witnessed the surge in popularity of the codebook model in the image domain. It was first introduced in text retrieval, and then later applied to image categorization in the seminal paper [40]. Subsequent research has focused on overcoming its two intrinsic limitations to improve discrimination, namely (1) the information loss of the assignment of local features to visual words, and (2) the lack of information on the spatial layout of the local features.

2.2.1 Quantization issues

Increasing the size of the dictionary is often reported to be able to improve the performance of the codebook model, but leads to a higher computational complexity for dictionary construction and feature assignment. On the other

hand, when the vocabularies are more compact, the information lost in the quantization process becomes more important, in particular when using hard assignment [41]. By directly using the image-to-class distances without descriptor quantization, Boiman et al. [42] showed that the discrimination ability is considerably decreased due to the rough quantization of the feature space. But with the soft-assignment of descriptors to multiple visual words, the loss can be compensated as reported in [43, 44]. Inspired by compressive sensing methodology, other approaches for assignment were guided by sparsity constraints [45] and locality constraints [46].

Bag-of-features (BoF) usually encodes the 0-order statistics of the distribution of descriptors. The Fisher vector extends the BoF by encoding high-order statistics (first and, optionally, second order). This description vector is the gradient of the sample's likelihood with respect to the parameters of this distribution, scaled by the inverse square root of the Fisher information matrix [47]. A simplified version of Fisher kernels, the vector of locally aggregated descriptors (VLAD) was also proposed in [48]. These three different ways of aggregating local image descriptors into a vector were evaluated by Jegou et al. [49]. Also, Picard and Gosselin [50] expanded the VLAD approach by adding an aggregation of the tensor product of descriptors.

In this paper, the description of 3D shapes is obtained by densely sampling salient points on the surface of the shape. In other words, the spectral signatures on every mesh vertex are considered to obtain the codebook representation. To shed some light on the feature space, we use the Laplacian and Gaussian kernels. We also use different kinds of ambiguity modeling methods to help us understand the information loss in quantization.

2.2.2 *Spatial information*

Similar to the image domain, the codebook model representation for 3D surfaces is a frequency histogram of quantized local geometric appearance, where the spatial layout of the geometric appearance is completely ignored [23]. Clearly, the spatial information may convey useful cues to improve the discrimination between 3D shapes. Before modeling the spatial layout on surfaces, it is necessary to review the technique for images. In the literature, two different ways to encode spatial information have been explored, which are based on local relative positions of pairwise features, and on global absolute positions.

2.2.3 *Relative spatial relation*

Modeling pairwise spatial features into the codebook model is an intuitive way to incorporate spatial information. A spatially sensitive affine-invariant image descriptor was constructed by Bronstein and Bronstein [51] using canonical relation, in which both the features and their relation are affine-invariant. They also generalize the pairwise spatially sensitive descriptors called "Expression" for 3D surface using the heat kernel as the relation [23]. Moreover, the relationship of visual words was also considered. Saverese et al. [52] used correlograms of visual words to model the spatial correlations between quantized local descriptors. Ling and Soatto [53] characterized the relative locations of visual words. Their proximity distribution representation is a 3D structure which records the number of times a visual word appears within a particular number of nearest neighbors of another word. Finally, besides pairwise relation, more complex relation such as the graph manner layout of groups of quantized local invariant descriptors was proposed by Behmo et al. [54], which can preserve translational relations between features. Liu et al. [55] calculated spatial histograms where the co-occurrences of local features are computed in circular regions of varying distances.

2.2.4 *Absolute spatial relation*

The spatial pyramid kernel (SPM), proposed by Lazebnik et al. [24], was one of the first works to address the lack of spatial information in the BoF representation. Their spatial pyramid representation was motivated by an earlier work, termed pyramid matching by Grauman and Darrell [56], on finding approximate correspondences between sets of points in high-dimensional feature spaces. The fundamental idea behind pyramid matching is to partition the feature space into a sequence of increasingly coarser grids and then compute a weighted sum over the number of matches that occur at each level of resolution. However, SPM and relative spatial relation modeling are still too weak. Recently, stronger spatially encoding methods include encoding geometric information of objects within the images. Local features of an image are projected onto different directions or points to generate a series of ordered BoF, based on which families of spatial partitions can guarantee the invariance of object translation, rotation, and scaling [57]. Additionally, there are some methods characterizing both the absolute and relative spatial layout of an image. Spatial pyramid co-occurrence [58] computes local co-occurrence with respect to spatial layout over a hierarchical spatial partitioning of an image. In addition to co-occurrences, geometry-preserving visual phrases [59] can encode more spatial information through capturing the local and long-range spatial layouts of the words. Unlike manually defined spatial regions for pooling, Jia et al. [60] proposed to learn more adaptive receptive fields to increase the performance even with a significantly smaller codebook size at the coding layer. In [61], the Gaussian mixture model was

encoded with spatial layout to improve the performance of Fisher kernel for image classification.

Local relative position was generalized to 3D surfaces as spatially sensitive descriptor in [23], but modeling the absolute spatial positions is not straightforward as in the case of 2D images because of the intrinsic deformation of nonrigid shapes. One contribution of this paper is to introduce the intrinsic partition to capture the global absolute spatial position, thus significantly improving the performance.

3 Deformation invariant shape representation

3.1 Laplace–Beltrami operator

Let $\mathbb{M}$ be a smooth orientable 2-manifold (surface) embedded in $\mathbb{R}^3$. A global parametric representation (embedding) of $\mathbb{M}$ is a smooth vector-valued map (also called surface patch) $\boldsymbol{x}$ defined from a connected open set (parametrization domain) $U \subset \mathbb{R}^2$ to $\mathbb{M} \subset \mathbb{R}^3$ such that

$$\boldsymbol{x}(\boldsymbol{u}) = \left(x^1(\boldsymbol{u}), x^2(\boldsymbol{u}), x^3(\boldsymbol{u})\right) \tag{1}$$

where $\boldsymbol{u} = (u^1, u^2) \in U$.

Given a twice-differentiable function $f : \mathbb{M} \to \mathbb{R}$, the LB operator [19] is a second-order partial differential operator defined as

$$\begin{aligned}\Delta_{\mathbb{M}} f &= -\frac{1}{\sqrt{|g|}} \sum_{i,j=1}^{2} \frac{\partial}{\partial u^j}\left(\sqrt{|g|} g^{ij} \frac{\partial f}{\partial u^i}\right) \\ &= -\sum_{i,j=1}^{2} g^{ij} \frac{\partial}{\partial u^j} \frac{\partial f}{\partial u^i} + (\text{lower order terms})\end{aligned} \tag{2}$$

where the matrix $g = (g_{ij})$ is referred to as a *Riemannian metric tensor* on $\mathbb{M}$, g^{ij} denote the elements of the inverse of the metric tensor g^{-1}, and $|g|$ is the determinant of g. The functions g_{ij} are sometimes referred to as the *metric coefficients*. The Riemannian metric g is an intrinsic quantity in the sense that it relates to measurements inside the surface. It is the analogous of the speed in the case of space curves, and determines all the intrinsic properties of the surface $\mathbb{M}$. These properties depend on the surface and do not depend on its embedding in space. In addition, the tensor g is invariant to rotation of the surface in space because it is defined in terms of inner products that are rotation invariant.

3.2 Discretization

Assume that the surface $\mathbb{M}$ is approximated by a triangular mesh. A triangle mesh $\mathbb{M}$ may be defined as $\mathbb{M} = (\mathcal{V}, \mathcal{E})$ or $\mathbb{M} = (\mathcal{V}, \mathcal{T})$, where $\mathcal{V} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$ is the set of vertices, $\mathcal{E} = \{e_{ij}\}$ is the set of edges, and $\mathcal{T} = \{\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n\}$ is the set of triangles. Each edge e_{ij} (denoted by $[\boldsymbol{v}_i, \boldsymbol{v}_j]$ or simply $[i, j]$) connects a pair of vertices $\{\boldsymbol{v}_i, \boldsymbol{v}_j\}$. Two distinct vertices $\boldsymbol{v}_i, \boldsymbol{v}_j \in \mathcal{V}$ are adjacent (denoted by $\boldsymbol{v}_i \sim \boldsymbol{v}_j$ or simply $i \sim j$) if they are connected by an edge, i.e., $e_{ij} \in \mathcal{E}$. The neighborhood (1-ring) of a vertex $\boldsymbol{v}_i$ is the set $\boldsymbol{v}_i^\star = \{\boldsymbol{v}_j \in \mathcal{V} : i \sim j\}$.

Several discretizations of the LB operator are available in the literature [16, 62–65]. In this paper, we use the approach developed in [62], which employs a mixed finite element/finite volume method on triangle meshes. Hence, the value of $\Delta_{\mathbb{M}} f$ at a vertex $\boldsymbol{v}_i$ can be approximated using the cotangent weight scheme as follows:

$$\Delta_{\mathbb{M}} f(\boldsymbol{v}_i) = \frac{1}{a_i} \sum_{\boldsymbol{v}_j \in \boldsymbol{v}_i^\star} \frac{\cot \alpha_{ij} + \cot \beta_{ij}}{2} \left[f(\boldsymbol{v}_j) - f(\boldsymbol{v}_i)\right], \tag{3}$$

where α_{ij} and β_{ij} are the angles $\angle(\boldsymbol{v}_i \boldsymbol{v}_{k_1} \boldsymbol{v}_j)$ and $\angle(\boldsymbol{v}_i \boldsymbol{v}_{k_2} \boldsymbol{v}_j)$ of two faces $\boldsymbol{t}^\alpha = \{\boldsymbol{v}_i, \boldsymbol{v}_j, \boldsymbol{v}_{k_1}\}$ and $\boldsymbol{t}^\beta = \{\boldsymbol{v}_i, \boldsymbol{v}_j, \boldsymbol{v}_{k_2}\}$ that are adjacent to the edge $[i, j]$, and a_i is the area of the voronoi cell (shaded polygon), as shown in Fig. 1. It is worth pointing out that the cotangent weight scheme is numerically consistent and preserves several important properties of the continuous LB operator, including symmetry and positive-definiteness [63].

Define the weight function $\omega : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ as

$$\omega_{ij} = \begin{cases} \dfrac{\cot \alpha_{ij} + \cot \beta_{ij}}{2a_i} & \text{if } i \sim j \\ 0 & \text{o.w.} \end{cases} \tag{4}$$

Then, for a function $f : \mathcal{V} \to \mathbb{R}$ that assigns to each vertex $\boldsymbol{v}_i \in \mathcal{V}$ a real value $f(\boldsymbol{v}_i)$ (we can view f as a column vector of length m), we may write the LB operator given by Eq. (3) as

$$Lf(\boldsymbol{v}_i) = \sum_{\boldsymbol{v}_j \in \boldsymbol{v}_i^\star} \omega_{ij} (f(\boldsymbol{v}_i) - f(\boldsymbol{v}_j)), \tag{5}$$

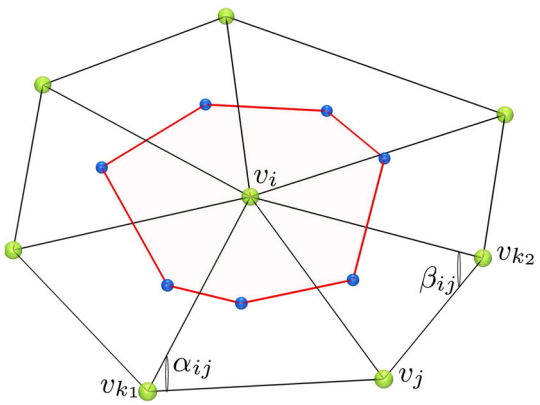


Fig. 1 Cotangent weight scheme: illustration of the angles α_{ij} and β_{ij}

where the matrix L is given by

$$L = \begin{cases} d_j & \text{if } i = j \\ -\omega_{ij} & \text{if } i \sim j \\ 0 & \text{o.w.} \end{cases} \tag{6}$$

and $d_j = \sum_{i=1}^{m} \omega_{ij}$ is the weighted degree of the vertex $\boldsymbol{v}_i$.

3.3 Eigenanalysis

Note that $\omega_{ij} \neq \omega_{ji}$ implies L is not a symmetric matrix. Thus, the spectrum (set of eigenvalues) of the eigenvalue problem $L\boldsymbol{\varphi}_i = \lambda_i \boldsymbol{\varphi}_i$ may not be real [15]. Noting that $\omega_{ij} = c_{ij}/a_i$, where

$$c_{ij} = \begin{cases} \dfrac{\cot \alpha_{ij} + \cot \beta_{ij}}{2} & \text{if } i \sim j \\ 0 & \text{o.w.} \end{cases} \tag{7}$$

we may factorize the matrix L as $L = A^{-1}C$, where $A = \text{diag}(a_i)$ is a positive-definite diagonal matrix and C is a sparse symmetric matrix given by

$$C = \begin{cases} \sum_{i=1}^{m} c_{ij} & \text{if } i = j \\ -c_{ij} & \text{if } i \sim j \\ 0 & \text{o.w.} \end{cases} \tag{8}$$

Therefore, we may write the eigenvalue problem $L\boldsymbol{\varphi}_i = \lambda_i \boldsymbol{\varphi}_i$ as a generalized eigenvalue problem $C\boldsymbol{\varphi}_i = \lambda_i A \boldsymbol{\varphi}_i$, which can be solved efficiently using the Arnoldi method of ARPACK. Figure 2 shows a 3D hand model and the sparsity patterns of the cotangent matrix C. Recall that the sparsity pattern (or support) of a matrix $A = (a_{ij})$ is the set of indices ij with $a_{ij} \neq 0$.

4 Graph embedding and shape descriptors

The eigenvalues λ_i and associated eigenfunctions $\boldsymbol{\varphi}_i$ of the LB operator can be computed by solving the generalized eigenvalue problem:

$$C\boldsymbol{\varphi}_i = \lambda_i A \boldsymbol{\varphi}_i, \quad i = 1, 2, \ldots, m \tag{9}$$

where $\boldsymbol{\varphi}_i$ is the unknown eigenfunction evaluated at m mesh vertices. That is, $\boldsymbol{\varphi}_i$ is an m-dimensional vector. We may sort the eigenvalues in ascending order as $0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_m$ with corresponding eigenfunctions as $\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \ldots, \boldsymbol{\varphi}_m$, where each eigenfunction $\boldsymbol{\varphi}_i = (\varphi_i(\boldsymbol{v}_1), \ldots, \varphi_i(\boldsymbol{v}_m))'$ is an m-dimensional vector. Note that the eigensystem $\{\lambda_i, \boldsymbol{\varphi}_i\}_i$ is intrinsic to the manifold and enjoys a nice property of being isometric invariant. It should also be noted that the meshes are assumed to be connected.

4.1 ShapeDNA and eigenvalue descriptors

The ShapeDNA [16] is one of the first spectral shape descriptors. It is a normalized sequence of the first eigenvalues of the LB operator. Its main advantages are the simple representation (a vector of numbers) and scale invariance. Despite its simplicity, the ShapeDNA yields a very good performance for the shape retrieval of nonrigid shapes. However, the eigenvalues are a global descriptor, therefore the ShapeDNA cannot be used for local or partial shape analysis. The Eigenvalue descriptor (EVD) [5], on the other hand, is a sequence of the eigenvalues of the geodesic distance matrix. Both ShapeDNA and EVD can be normalized by the second eigenvalue.

4.2 Global point signature

The GPS [15] at a surface point is a vector of scaled eigenfunctions of the LB operator. The GPS is a global feature in the sense that it cannot be used for partial shape matching. It is defined in terms of the eigenvalues and eigenfunctions of $\Delta_{\mathbb{M}}$ as follows:

$$\text{GPS}(x) = \left(\frac{\varphi_2(x)}{\sqrt{\lambda_2}}, \frac{\varphi_3(x)}{\sqrt{\lambda_3}}, \ldots, \frac{\varphi_i(x)}{\sqrt{\lambda_i}}, \ldots \right) \tag{10}$$

GPS is invariant under isometric deformations of the shape, but it suffers for the problem of eigenfunctions switching whenever the associated eigenvalues are close to each other.

4.3 Heat kernel signature

The heat kernel $\mathfrak{p}_t(x, y)$ is a fundamental solution to the heat equation [66] at point x at time t with initial distribution $u_0(x) = \delta(x - y)$ at point $y \in \mathbb{M}$, and it is defined in terms of the eigenvalues and eigenfunctions of $\Delta_{\mathbb{M}}$ as follows:

$$\mathfrak{p}_t(x, y) = \sum_{i=1}^{\infty} e^{-\lambda_i t} \varphi_i(x) \varphi_i(y) \tag{11}$$

Intuitively, $\mathfrak{p}_t(x, y)$ describes the amount of heat that is propagated or transferred from point x to point y in time t. Therefore, $\mathfrak{p}_t(x, x)$ describes the amount of heat that remains at point x after time t. For each point $x \in \mathbb{M}$, the HKS [11] is represented in the discrete temporal domain by a n-dimensional feature vector

$$\text{HKS}(x) = \left(\mathfrak{p}_{t_1}(x, x), \mathfrak{p}_{t_2}(x, x), \ldots, \mathfrak{p}_{t_n}(x, x) \right) \tag{12}$$

where $t_1, t_2, \ldots, t_n$ are different time-scales.

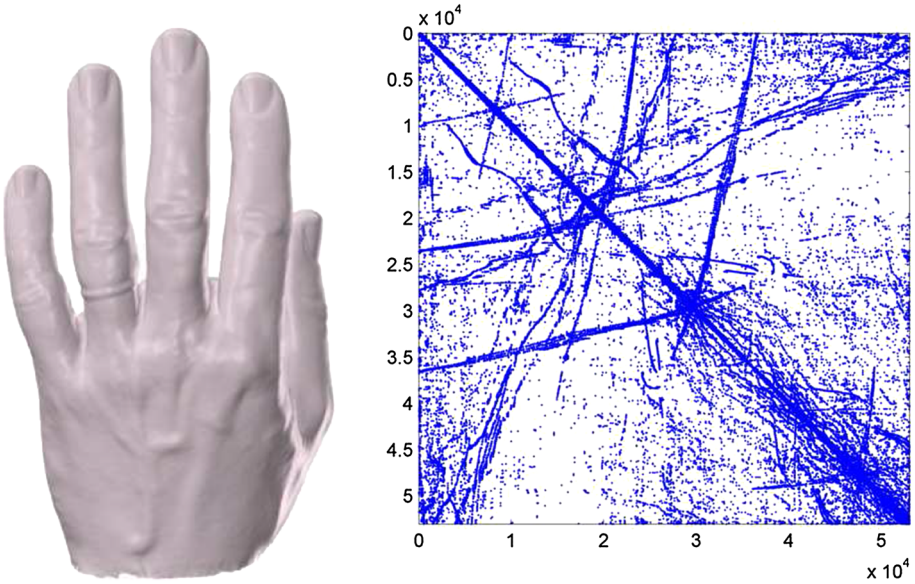


Fig. 2 3D hand model (*left*) and sparsity pattern plot of the cotangent matrix C (*right*)

4.4 Scale invariant heat kernel signature

Let $\mathbb{M}$ and $\mathbb{M}'$ be a shape and its uniformly scaled version by a factor of a, respectively. Denote by $\mathfrak{p}_{\alpha^\tau}(x,y)$ the heat kernel with scale logarithmically sampled using some basis α at each point x. Thus, the heat kernel of the scaled shape becomes $\mathfrak{p}'(\tau) = a^{-2}\mathfrak{p}(\tau + 2\log_\alpha a)$. In order to remove the dependence on the multiplicative constant a^{-2}, the logarithm of the signal is taken and then differentiated with respect to the scale variable [12]:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\tau}\log \mathfrak{p}'(\tau) &= \frac{\mathrm{d}}{\mathrm{d}\tau}(-2\log a + \log \mathfrak{p}(\tau + 2\log_\alpha a)\\ &= \frac{\frac{\mathrm{d}}{\mathrm{d}\tau}\mathfrak{p}(\tau + 2\log_\alpha a)}{\mathfrak{p}(\tau + 2\log_\alpha a)}\end{aligned} \tag{13}$$

Let $\mathfrak{p}' = \frac{\frac{\mathrm{d}}{\mathrm{d}\tau}\mathfrak{p}(\tau)}{\mathfrak{p}(\tau)} = \frac{-\sum_{i\geq 0}\lambda_i\alpha^\tau \log\alpha e^{-\lambda_i\alpha^\tau}\varphi_i^2(x)}{-\sum_{i\geq 0}e^{-\lambda_i\alpha^\tau}\varphi_i^2(x)}$ then a new function $\tilde{\mathfrak{p}}$ which transforms $\tilde{\mathfrak{p}}'(\tau) = \tilde{\mathfrak{p}}(\tau + 2\log_\alpha a)$ as a result of scaling is obtained. The Fourier transform of $\tilde{\mathfrak{p}}$ and its absolute value are given by

$$\begin{aligned}&F[\tilde{\mathfrak{p}}'](\omega) = \tilde{H}'(\omega) = \tilde{H}(\omega)e^{-j\omega 2\log_\alpha a}\\ &|\tilde{H}'(\omega)| = |\tilde{H}(\omega)|.\end{aligned} \tag{14}$$

Thus, the SIHKS is defined as

$$\mathrm{SIHKS}(x) = \big(|\tilde{H}(\omega_1)|, |\tilde{H}(\omega_2)|, \ldots, |\tilde{H}(\omega_n)|\big). \tag{15}$$

4.5 Wave kernel signature

The fundamental idea of the WKS [14] is to represent a point $x \in \mathbb{M}$ by the average probabilities of quantum particles of different energy levels to be measured in x. Assume a quantum particle with unknown position is on the surface. Then the wave function of the particle is the Schrödinger equation solution, which can expressed in the spectral domain as

$$\psi_E(x,t) = \sum_{k=1}^{\infty} e^{i\lambda_k t}\varphi_k(x)f_E(\lambda_k) \tag{16}$$

where E denotes the energy of the particle at time $t = 0$ and f_E its initial distribution.

Since $|\psi_E(x,t)|^2$ is the probability to measure the particle at a point x at time t, it follows that the average probability (over time) to measure a particle in x is given by

$$\mathrm{P}_E(x) = \lim_{T\to\infty}\frac{1}{T}\int_0^T |\psi_E(x,t)|^2 = \sum_{k=1}^{\infty}\varphi_k(x)^2 f_E(\lambda_k)^2 \tag{17}$$

Let $E_1, E_2, \ldots, E_n$ be n log-normal energy distributions. Then, each point x on the surface $\mathbb{M}$ is associated with a WKS, which can represented by a n-dimensional feature vector of average probabilities as follows:

$$\mathrm{WKS}(x) = (\mathrm{P}_{e_1}(x), \mathrm{P}_{e_2}(x), \ldots, \mathrm{P}_{e_n}(x)) \tag{18}$$

where $e_i = \log E_i$ is the logarithmic energy scale. The WKS represents the average probability of measuring a quantum particle at a specific surface point. Unlike the HKS, the WKS separates influences of different frequencies, treating all frequencies equally. In other words, HKS uses low-pass filters, while WKS uses band-pass filters.

4.6 Heat mean signature

The HMS [13] quantitatively evaluate the temperature distribution resulting from the heat flow process

$$\mathrm{HMS}_t(x) = \frac{1}{m}\sum_{y \neq x} \mathfrak{p}_t(x,y) \tag{19}$$

which can be physically interpreted as the average temperature on the surface obtained by applying a unit amount of heat on the vertex x and after a certain amount of time of heat dissipation. A relatively smaller parameter t is often empirically chosen to preserve a higher resolution version of the original surface [67]. Fang et al. [17] also proposed the temperature distribution descriptor, which is based on the distribution of the values of average temperature for all of the vertices on the mesh. We construct a multi-scale HMS to compare temperature distribution with multiple diffusion times as follows:

$$\mathrm{HMS}(x) = (\mathrm{HMS}_{t_1}, \mathrm{HMS}_{t_2}, \ldots, \mathrm{HMS}_{t_n}). \tag{20}$$

For the sake of notational simplicity, we use $s(x)$ to represent the types of the above spectral signatures evaluated at a surface point x, i.e., GPS, HKS, SIHKS, WKS, or HMS.

5 Aggregating local descriptors with codebook models

In this section, we focus on the methods for aggregating dense local spectral descriptors into a compact representation of the whole shape. Bronstein et al. [23] initiated the study of the nonrigid 3D shape retrieval via BoF with soft-assignment. In this paper, we comprehensively investigate some variants of the codebook model for aggregating these local spectral descriptors in a dense signature space. In particular, we propose an intrinsic spatial partition, which can be seen as the counterpart of the spatial extension for the codebook model in image recognition; thus further enhancing the results.

Given a set of local point-wise signatures densely computed on each vertex on the mesh surface, we quantize the signature space to obtain a compact histogram representation of the shape using the codebook model approach. The visual word vocabulary in the codebook model may be constructed in various ways. We use the k-means algorithm to generate the visual vocabulary. This is computationally expensive, but as this step is performed off-line, it has no impact on the search time. Thus, the "geometric words" of a vocabulary $P = \{p_k, k = 1, 2, \ldots, K\}$ are obtained as the K centroid of k-means clustering in the signature space. We can use various types of spectral descriptors. From any shape, we get a specific type of local spectral descriptors $S = \{s_t, t = 1, 2, \ldots, T\}$ for comparison. By a certain vector coding technique, each shape will be described by a histogram H. Since the number of vertices is usually different among different meshed shapes, an appropriate normalization technique is also important for the codeword-cumulative histogram representation. We normalize P by the total number of vertices of each shape.

Modeling the codeword ambiguity plays a crucial role on the performance of the codebook model. In the literature, visual word ambiguity modeling is used occasionally, often ad hoc motivated, and rarely evaluated. However, a formal summarized work was recently proposed by Gemert et al. [44], who motivated and evaluated several types of visual word ambiguity, and provided ample analysis. For completeness, we introduce these types in the scenario of 3D geometric shapes:

- L_p **norm codebook:** Each local descriptor s_t is associated with its nearest visual word $NN(s_t)$ in the codebook. For each codeword p_k, the differences of vector s_t assigned to p_k are accumulated by L_p norm as follows:

$$q_i = \sum_{s_t : NN(s_t) = i} \|s_t - p_i\|_\mathrm{p}, \quad \mathrm{p} = 0, 1, 2 \tag{21}$$

  Note that L_0 is the traditional codebook, which is the histogram of the number of local descriptors assigned to each visual word.
- **Kernel codebook:** The histogram estimator of the code words may be replaced by a kernel density estimator. Moreover, a suitable kernel (such as the Gaussian) allows kernel density estimation to become part of the codewords, rather than the data samples. A symmetric kernel allows for transferring the kernel from the data samples to the codewords, yielding a kernel codebook

$$q_i = \sum_{t=1}^{T} K_\sigma(\varrho(s_t, p_i)), \tag{22}$$

  In order to make it clear which distance is more fit for the descriptors, we use both the L_1 and L_2 norms as distance functions. The Euclidean distance (L_2-norm) is paired with a Gaussian-shaped kernel, while the L_1-norm is paired with a Laplacian-shaped kernel. The latter assumes that the variation between a local feature and a codeword is described by a sharper distribution. Both distributions have a smoothing parameter σ which represents the size of the kernel. For simplicity, we denote the kernels as $K_\sigma(\varrho(\cdot,\cdot))$, where $\varrho(\cdot,\cdot)$ is the L_1-norm when it is the Laplacian kernel, and L_2-norm when it is the Gaussian kernel.
- **Codeword uncertainty:** Codeword uncertainty indicates that one image region may distribute probability mass to more than one codeword. It is modeled to normalize the amount of probability mass to a total constant weight of 1 and is distributed over all relevant

codewords. Relevancy is determined by the ratio of the kernel values for all codewords p_i in the vocabulary

$$q_i = \sum_{t=1}^{T} \frac{K_\sigma(\varrho(s_t, p_i))}{\sum_{k=1}^{K} K_\sigma(\varrho(s_t, p_k))} \tag{23}$$

- **Codeword plausibility:** Codeword plausibility means that an image feature may not be close enough to warrant representation by any relevant codeword in the vocabulary. For a given descriptor s_t, it selects the best fitting codeword p_i and assigns its probability mass proportional to the kernel value of that codeword. Hence, codeword plausibility will give a higher weight to more relevant data samples. However, it cannot select multiple codeword candidates

$$q_i = \sum_{s_t : NN(s_t) = i} K_\sigma(\varrho(s_t, p_i)). \tag{24}$$

The four types of ambiguity modeling methods use different numbers of geometric words in their construction. In the traditional codebook and codeword plausibility, the local descriptor only selects the best candidate geometric word. On the other hand, the kernel codebook and codeword uncertainty divide the descriptor over multiple best fitting codewords. To formally compare the different ambiguity ways, we summarize all the codebook models in Table 1 and we categorize them in terms of the L_1- and L_2-norms.

The kernel size determines the degree of coherence to assign geometric word to a descriptor, and it is dependent on the descriptor dimensionality and the range of the descriptor values. Moreover, we only consider the kernel size that is fixed for all codewords. The case of constructing the variable kernel density estimator for different codewords can also be considered, but we adhere to a homogenous feature space by keeping the kernel size fixed for all codewords [44]. Note that we do not try to obtain the best fit of the data. In contrast, we aim at finding the kernel size that discriminates well between classes. In the experimental results section, we estimate the optimal kernel size in an interval inferred from the data distribution.

Besides directly modeling ambiguity on individual geometric words, ambiguity might be addressed by modeling geometric word co-occurrences. Co-occurrence modeling may address ambiguity because it is likely that similar geometric words with high ambiguity co-occur frequently. When these ambiguous geometric words are grouped together, their intra-ambiguity is resolved. For 3D shapes, Bronstein and Bronstein [23] introduced the spatially sensitive Bag-of-Words description, which accounts not only for the frequency but also for the spatial relations between features. In this paper, since we are interested in analyzing different spectral descriptors and measuring ambiguity, we concentrate on single word ambiguity modeling.

6 Intrinsic spatial partition: beyond codebook model

6.1 Isocontours

The eigenfunctions of the LB operator enjoy nice properties including isometry invariance and robustness to pose

Table 1 Codebook model

Clustering norm	Vector assignment	
L_1	Traditional codebook	$q_i = \frac{1}{T}\sum_{s_t:NN(s_t)} \|s_t - p_i\|_0$
	L_1-norm codebook	$q_i = \frac{1}{T}\sum_{s_t:NN(s_t)} \|s_t - p_i\|_1$
	L_2-norm codebook	$q_i = \frac{1}{T}\sum_{s_t:NN(s_t)} \|s_t - p_i\|_2$
	Laplace kernel codebook	$q_i = \frac{1}{T}\sum_{t=1}^{T} \frac{1}{2b} e^{-\|t-i\|_1}$
	Laplace codeword uncertainty	$q_i = \frac{1}{T}\sum_{t=1}^{T} \frac{\frac{1}{2b}e^{-\|t-i\|_1}}{\sum_k \frac{1}{2}e^{-\|t-k\|_1}}$
	Laplace codeword plausibility	$q_i = \frac{1}{T}\sum_{s_t:NN(s_t)=i} \frac{1}{2b} e^{-\|t-i\|_1}$
L_1	Traditional codebook	$q_i = \frac{1}{T}\sum_{s_t:NN(s_t)} \|s_t - p_i\|_0$
	L_1-norm codebook	$q_i = \frac{1}{T}\sum_{s_t:NN(s_t)} \|s_t - p_i\|_1$
	L_2-norm codebook	$q_i = \frac{1}{T}\sum_{s_t:NN(s_t)} \|s_t - p_i\|_2$
	Gaussian kernel codebook	$q_i = \frac{1}{T}\sum_{t=1}^{T} \frac{1}{\sigma\sqrt{2}} e^{-\frac{\|t-i\|_2^2}{2\sigma^2}}$
	Gaussian codeword uncertainty	$q_i = \frac{1}{T}\sum_{t=1}^{T} \frac{\frac{1}{\sqrt{2}}e^{-\frac{1}{2}\|t-i\|_2^2}}{\sum_k \frac{1}{\sqrt{2}}e^{-\frac{1}{2}\|t-k\|_2^2}}$
	Gaussian codeword plausibility	$q_i = \frac{1}{T}\sum_{s_t:NN(s_t)=i} \frac{1}{\sigma\sqrt{2}} e^{-\frac{\|t-i\|_2^2}{2\sigma^2}}$

variations such as translation and rotation. These eigenfunctions are orthogonal $\langle \boldsymbol{\varphi}_i, \boldsymbol{\varphi}_j \rangle_A = 0, \forall i \neq j$, where the orthogonality is defined in terms of the A-inner product. That is, $\langle \boldsymbol{\varphi}_i, \boldsymbol{\varphi}_j \rangle_A = \boldsymbol{\varphi}_i' A \boldsymbol{\varphi}_j$. Moreover, any function $f : \mathcal{V} \to \mathbb{R}$ (viewed as a column-vector of length m) on the triangle mesh $\mathbb{M}$ can be written in terms of the eigenfunctions as follows:

$$f = \sum_{i=1}^{m} \alpha_i \boldsymbol{\varphi}_i, \quad \text{where} \quad \alpha_i = \langle f, \boldsymbol{\varphi}_i \rangle. \tag{25}$$

Note that since the sum of each row in the matrix C equals zero, the first eigenvalue λ_1 is zero and the corresponding eigenfunction $\boldsymbol{\varphi}_1$ is a constant m-dimensional vector. The top row of Fig. 3 shows a 3D horse model colored by the second, third and fourth eigenfunctions, while the bottom row displays the isocontours of these eigenfunctions.

We can use the variational characterizations of the eigenvalues in terms of the Rayleigh-Ritz quotient. That is, the second eigenvalue is given by

$$\lambda_2 = \inf_{f \perp \boldsymbol{\varphi}_1} \frac{f' C f}{f' A f} = \inf_{f \perp \boldsymbol{\varphi}_1} \frac{\sum_{i \sim j} c_{ij} (f(\boldsymbol{v}_i) - f(\boldsymbol{v}_j))^2}{\sum_i f(\boldsymbol{v}_i)^2 a_i} \tag{26}$$

and $\boldsymbol{\varphi}_2 = (\varphi_2(\boldsymbol{v}_1), \ldots, \varphi_2(\boldsymbol{v}_m))'$ is its corresponding eigenvector.

The eigenvalues and eigenfunctions have a nice physical interpretation: the square roots of the eigenvalues $\sqrt{\lambda_i}$ are the eigenfrequencies of the membrane, and $\varphi_i(x)$ are the corresponding amplitudes at x. In particular, the second eigenvalue λ_2 corresponds to the sound we hear the best. On the other hand, Uhlenbeck [68] showed that the eigenfunctions of the LB operator are Morse functions on the interior of the domain of the operator. Consequently, this generic property of the eigenfunctions gives rise to constructing the associated intrinsic isocurves.

6.2 Intrinsic spatial partition

Motivated by the isometric invariance property of the second eigenfunction of the LB operator and also by its generic property as a Morse function, we propose to use the level sets (isocontours) of the second eigenfunction as cuts to partition the surfaces. In Fig. 4a–c, we show some examples of the level curves of $\boldsymbol{\varphi}_2$. In Fig. 4a, we can observe that the isocontours are consistent with global large deformation (first column), local small bend (second column), and among the shapes from different classes, but share similar topological structure (third column). The correspondence of isocontours on the shapes from the same class are displayed in Fig. 4b, which shows models that include various topological structures. Finally, the consistency of isocontours on the shapes from the different class are displayed in Fig. 4c. Although the shapes are explicitly different, their isocontours can capture their intrinsic correspondence well.

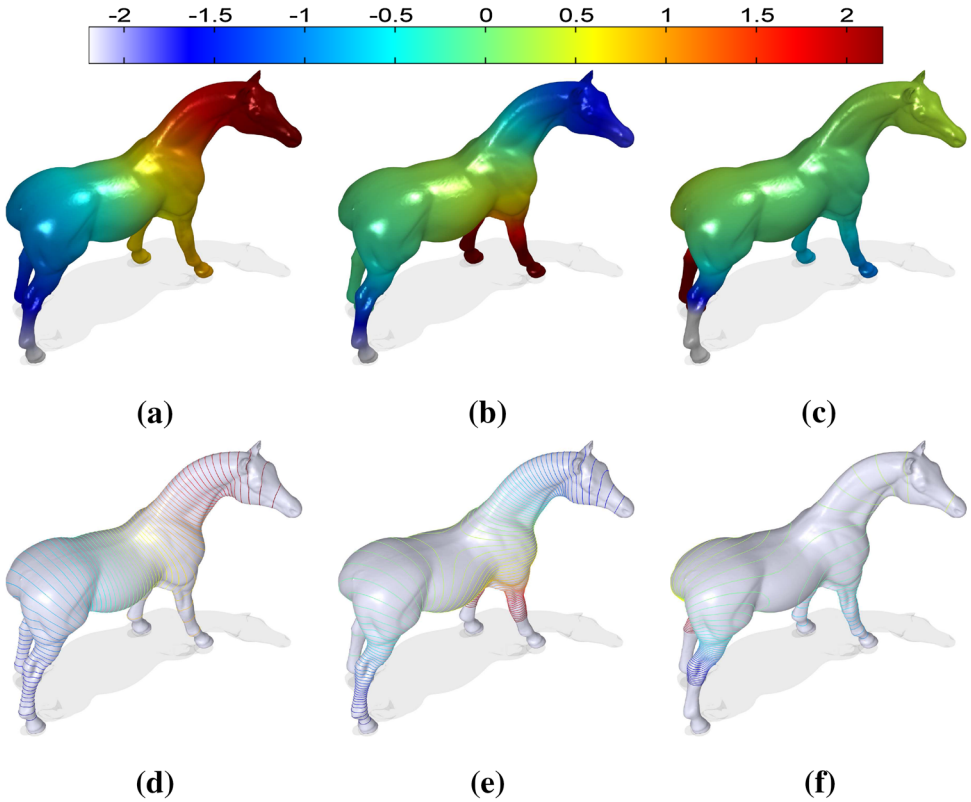


Fig. 3 **a–c** 3D horse model colored by $\varphi_2, \varphi_3, \varphi_4$. **d–f** Level sets of $\varphi_2, \varphi_3, \varphi_4$

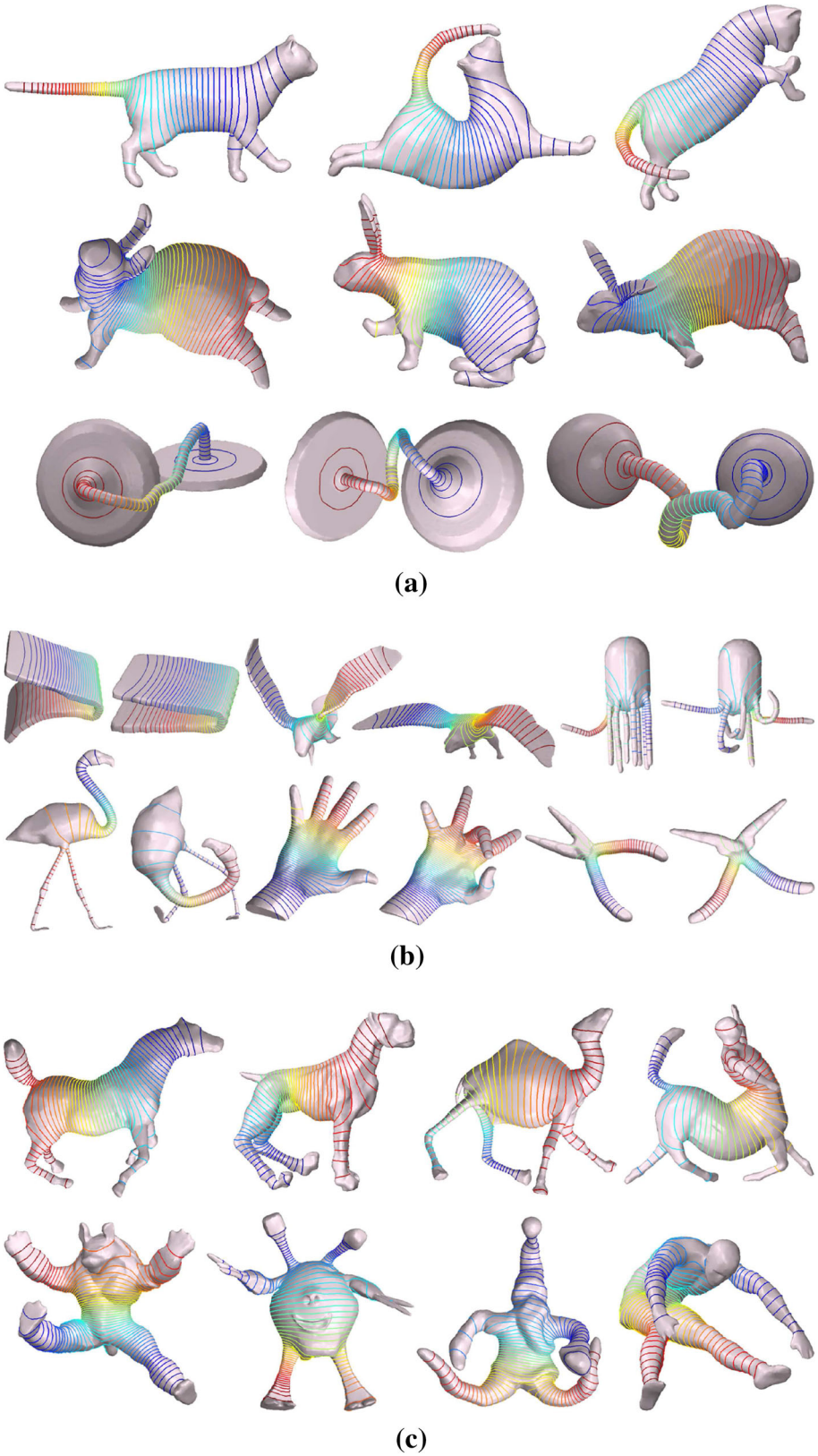


Fig. 4 **a** Isocontours are invariant under both global and local deformation. **b** Proportionality correspondence of pairwise nonrigid shapes with varied topological structure. **c** Isocontours are consistent among different classes of shapes

The level sets of the second eigenfunction have been used to extract curve skeletons of the nonrigid shapes [10, 25], which is a vivid clue that these isocontours capture the global topological structure of shapes. We visualize the procedure for extracting the curve skeleton in Fig. 5.

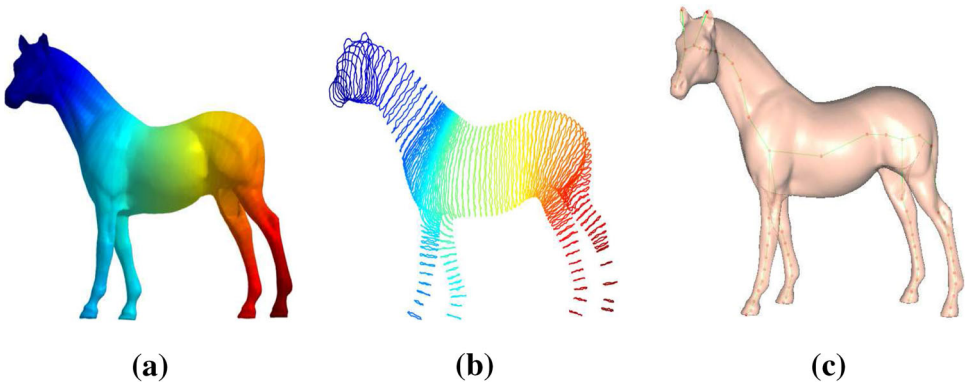


Fig. 5 **a** 3D horse model colored by φ_2; **b** level sets of φ_2; **c** spectral Reeb graph

Table 2 Parameters' setting for kernel size interval estimation on SHREC 2010 and SHREC 2011 datasets

	SHREC 2010					SHREC 2011				
	HKS	SIHKS	HMS	WKS	GPS	HKS	SIHKS	HMS	WKS	GPS
A	1	4	2	4	4	1	4	2	20	20
F	*A*	*A*	*A*	1	1	*A*	*A*	*A*	A/4	A/2

Table 3 Runtime (in seconds) with different descriptor dimensions and vocabulary sizes

Runtime	Vocabulary size							
Dimension	8	12	16	24	32	48	64	80
40	190	254	321	474	567	732	903	1,351
100	522	617	774	1,094	1,461	1,803	2,054	2,665
150	825	909	1,193	1,691	2,002	2,902	3,645	4,358
385	1,725	2,566	3,347	4,638	5,702	8,405	12,285	15,962

6.3 Matching by intrinsic spatial partition

Instead of representing the whole shape by the codeword model without considering spatial layout of local descriptors, we enhance the discrimination by integrating the distribution of local descriptors in different spatial patches determined by the intrinsic spatial partition. For any shape cut by isocontours at resolution R, its description H is the concatenation of R sub-histograms:

$$H = [h^1, h^2, \ldots, h^i, \ldots, h^{\mathrm{R}}] \tag{27}$$

where h^i is the sub-histogram ordered in the ith position according to the intrinsic spatial partition from one end to the other. Note that the isocontour sequence might start from either end, and the situations are different from shape to shape. For example, in Fig. 4a, the heads of the first and third rabbit are colored in blue, but tail of the second is colored in red and head in blue, whose order is exactly the opposite. To guarantee that the semantic correspondent parts are matched in the comparison, we use an order-insensitive strategy comparison method. First, we get a new histogram T by making the order of the sub-histogram inverted in H:

$$T = [h^{\mathrm{R}}, h^{(\mathrm{R}-1)}, \ldots, h^i, \ldots, h^1]. \tag{28}$$

Second, to compare two shapes P and Q we define their dissimilarity under this feature as follows:

$$\mathcal{B}^{\mathrm{R}}(P, Q) = \min(\mathcal{A}^{\mathrm{R}}(H_P, H_Q), \mathcal{A}^{\mathrm{R}}(H_P, T_Q)) \tag{29}$$

where H_P and H_Q denote the histograms of P and Q, respectively. In other words, there are two possible matching schemes between the isocontour sequences of two shapes, head-to-head and head-to-end. We consider the schemes with the minimum cost to be better matched. For each scheme, the dissimilarity measure $\mathcal{A}^{\mathrm{R}}(\cdot,\cdot)$ is defined as

$$\mathcal{A}^{\mathrm{R}}(H_P, H_Q) = \sum_{i=1}^{\mathrm{R}} \sum_{k=1}^{K} \Psi(h_P^i(k), h_Q^i(k)) \tag{30}$$

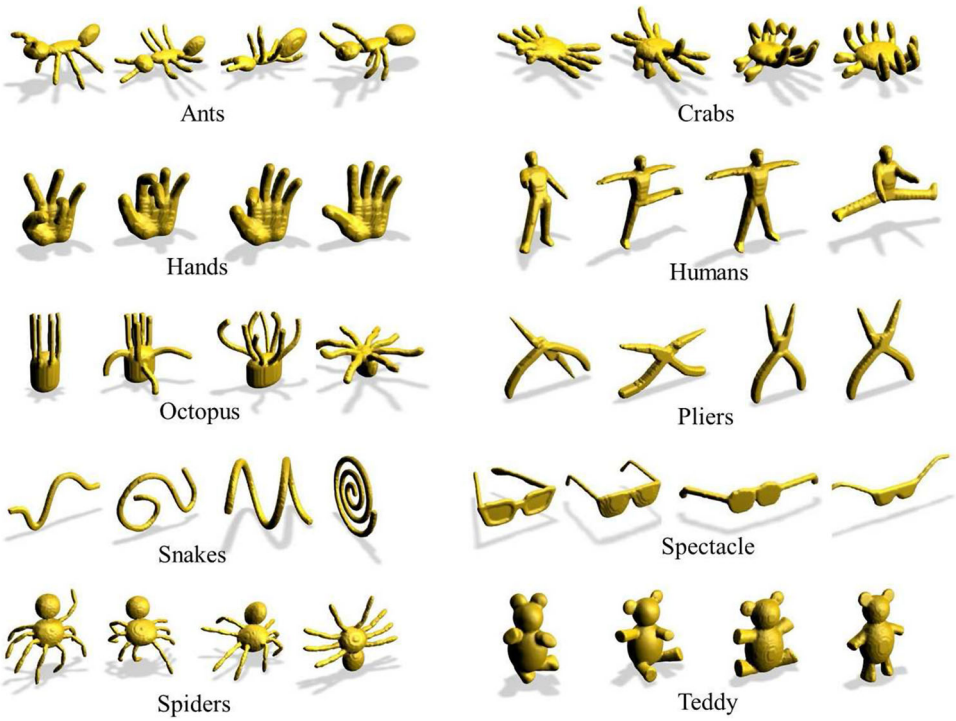


Fig. 6 Sample shapes in SHREC 2010 dataset

where $\Psi(\cdot,\cdot)$ can be any histogram comparison metric. In this paper, we use the chi-squared kernel so that $h_P^i(k)$ and $h_Q^i(k)$ are the accumulations of the code of the local descriptors from P and Q that fall into the kth codeword cell/channel of the ith patch.

The degree of resolution would affect the performance of the spatial partition-based method. To further improve the results, we extend the spatial pyramid [24], which has been shown to yield excellent performance in image analysis, to nonrigid 3D shapes. The spatial pyramid divides an image into a multi-level pyramid of increasingly fine subregions and computes a codebook descriptor for each subregion. We construct a sequence of histograms at resolutions $\{\mathtt{R} = 2^\ell, \ell = 0, \ldots, L\}$ such that the surface at level ℓ has 2^ℓ patches, for a total of $2^L - 1$ patches. Thus, the final dissimilarity between the histograms of P and Q is given by

$$\begin{aligned}\mathcal{D}^L(P,Q) &= \mathcal{B}^L(P,Q) + \sum_{\ell=0}^{L-1} \frac{1}{2^{L-\ell}} \left(\mathcal{B}^\ell(P,Q) - \mathcal{B}^{\ell+1}(P,Q)\right) \\ &= \frac{1}{2^L}\mathcal{B}^0(P,Q) + \sum_{\ell=1}^{L} \frac{1}{2^{L-\ell+1}} \mathcal{B}^\ell(P,Q) \end{aligned} \tag{31}$$

Concerning the implementation, one issue that arises is that of normalization. To easily compare the methods of single level partition and intrinsic spatial pyramid matching, we normalize the histogram of each resolution using the L_1-norm.

7 Experiments

We experimentally compare different spectral descriptors and the codeword ambiguity modeling approaches on two standard datasets: SHREC 2010 [70] and SHREC 2011 [71]. We also show that the proposed intrinsic partition approach can significantly improve the performance of the spectral shape retrieval methods. We start our experiments with an in-depth analysis of the methods on a set of ten 3D nonrigid shape categories, after which we translate these findings into the experiments on a large dataset. In our experimental setup, we closely follow the original works theoretically, and we select the optimal parameters that yield the best performance on these datasets. Thus, we do not bias any method in order to provide a fair comparison.

7.1 Settings

7.1.1 Evaluation measure

We evaluate the retrieval performance using the discounted cumulative gain (DCG) [69]. DCG is a statistic that weights correct results near the front of the list more

Table 4 DCG values for different spectral signatures and classes on SHREC 2010

Signature	Ant	Crab	Hand	Human	Octopus	Plier	Snake	Spectacle	Spider	Teddy
SIHKS	0.951	0.901	0.897	0.884	0.783	0.935	0.714	0.773	0.899	0.990
HKS	0.925	0.877	0.804	0.805	0.726	0.979	0.707	0.732	0.886	0.990
WKS	0.793	0.813	0.652	0.704	0.731	0.784	0.713	0.744	0.723	0.913
HMS	0.668	0.785	0.709	0.687	0.646	0.880	0.696	0.633	0.731	0.998
GPS	0.846	0.675	0.713	0.744	0.576	0.881	0.622	0.675	0.727	0.821

Table 5 Performance (DCG) using different codebook models of varying size based on SIHKS local descriptor

Codebook model		Vocabulary size								
Clustering	Coding	8	12	16	24	32	48	64	80	200
L_1	Traditional codebook	0.801	0.789	0.777	0.779	0.798	0.795	0.802	0.795	0.793
	L_1-norm codebook	0.801	0.796	0.784	0.789	0.806	0.801	**0.813**	0.806	0.803
	L_2-norm Codebook	0.776	0.768	0.761	0.770	0.791	0.785	0.804	0.792	0.792
	Laplace kernel codebook	0.797	0.795	0.796	0.796	0.798	0.801	0.803	0.810	0.812
	Laplace codeword uncertainty	0.809	0.803	0.805	0.802	0.808	0.810	0.811	0.812	0.806
	Laplace codeword plausibility	0.798	0.782	0.774	0.774	0.792	0.792	0.798	0.793	0.783
L_2	Traditional codebook	0.856	0.863	0.861	0.862	0.872	0.865	0.865	0.866	0.849
	L_1-norm codebook	0.839	0.852	0.853	0.855	0.865	0.865	0.863	0.865	0.854
	L_2-norm codebook	0.841	0.846	0.852	0.849	0.854	0.861	0.858	0.861	0.846
	Gaussian kernel codebook	0.847	0.839	0.840	0.842	0.850	0.846	0.845	0.847	0.831
	Gaussian codeword uncertainty	0.857	0.863	0.864	0.867	0.874	0.867	0.869	0.867	0.827
	Gaussian codeword plausibility	0.856	0.863	0.861	0.863	0.872	0.865	0.866	0.866	0.841

Bold value indicates the best result

Table 6 Performance (DCG) using different codebook models of varying size based on HMS local descriptor

Codebook model		Vocabulary size								
Clustering	Coding	8	12	16	24	32	48	64	80	200
L_1	Traditional codebook	0.683	0.717	0.729	0.749	0.747	0.733	0.745	0.746	0.725
	L_1-norm codebook	0.693	0.721	0.752	0.760	0.759	0.711	0.766	0.751	0.723
	L_2-norm codebook	0.680	0.695	0.695	0.698	0.702	0.675	0.726	0.728	0.673
	Laplace kernel codebook	0.689	0.705	0.711	0.720	0.718	0.726	0.719	0.724	0.740
	Laplace codeword uncertainty	0.691	0.722	0.736	0.751	0.754	0.735	0.755	0.745	0.741
	Laplace codeword plausibility	0.680	0.709	0.719	0.738	0.740	0.733	0.734	0.740	0.724
L_2	Traditional codebook	0.696	0.726	0.737	0.746	0.743	0.743	0.749	0.753	0.748
	L_1-norm codebook	0.713	0.743	0.763	0.766	0.771	0.740	0.773	0.780	0.740
	L_2-norm codebook	0.712	0.730	0.752	0.763	0.761	0.718	0.765	0.759	0.727
	Gaussian kernel codebook	0.699	0.721	0.731	0.732	0.732	0.733	0.732	0.724	0.722
	Gaussian codeword uncertainty	0.707	0.740	0.759	0.768	0.771	0.753	0.767	0.771	0.733
	Gaussian codeword plausibility	0.696	0.726	0.737	0.746	0.743	0.744	0.749	0.753	0.748

than correct results later in the ranked list, under the assumption that a user is less likely to consider elements near the end of the list. For a given query shape, DCG is defined as

$$\mathrm{DCG} = \frac{\mathrm{DCG}_N}{1 + \sum_{k=2}^{\tau} \frac{1}{\log_2(k)}} \tag{32}$$

where N is the total number of shapes in the database, τ is the number of shapes in the query's class (i.e., size of the class), and DCG_i is computed recursively as follows:

$$\mathrm{DCG}_i = \begin{cases} G_i & \text{if } i = 1 \\ \mathrm{DCG}_{i-1} + \dfrac{G_i}{\log_2(i)} & \text{Otherwise} \end{cases} \tag{33}$$

Table 7 Performance (DCG) using different codebook models of varying size based on WKS local descriptor

Codebook model		Vocabulary size								
Clustering	Coding	8	12	16	24	32	48	64	80	200
L_1	Traditional codebook	0.691	0.711	0.720	0.727	0.726	0.741	0.737	0.737	0.748
	L_1-norm codebook	0.697	0.718	0.726	0.733	0.730	0.742	0.737	0.737	0.751
	L_2-norm codebook	0.694	0.711	0.712	0.721	0.719	0.731	0.725	0.725	0.733
	Laplace kernel codebook	0.694	0.701	0.707	0.713	0.714	0.725	0.724	0.726	0.738
	Laplace codeword uncertainty	0.697	0.716	0.721	0.725	0.727	0.740	0.736	0.737	0.741
	Laplace codeword plausibility	0.695	0.706	0.718	0.724	0.723	0.740	0.736	0.737	0.739
L_2	Traditional codebook	0.659	0.689	0.703	0.718	0.728	0.722	0.716	0.718	0.743
	L_1-norm codebook	0.666	0.704	0.710	0.724	0.730	0.728	0.721	0.722	0.738
	L_2-norm codebook	0.672	0.701	0.704	0.716	0.727	0.727	0.720	0.719	0.731
	Gaussian kernel codebook	0.666	0.667	0.667	0.666	0.666	0.666	0.667	0.667	0.651
	Gaussian codeword uncertainty	0.667	0.666	0.665	0.666	0.665	0.666	0.667	0.667	0.651
	Gaussian codeword plausibility	0.662	0.689	0.703	0.718	0.728	0.722	0.716	0.718	0.736

Table 8 Performance (DCG) using different codebook models of varying size based on GPS local descriptor

Codebook model		Vocabulary size								
Clustering	Coding	8	12	16	24	32	48	64	80	200
L_1	Traditional codebook	0.719	0.757	0.733	0.723	0.720	0.737	0.714	0.726	0.739
	L_1-norm codebook	0.716	0.752	0.729	0.720	0.717	0.732	0.708	0.723	0.738
	L_2-norm codebook	0.689	0.734	0.705	0.699	0.692	0.717	0.691	0.704	0.711
	Laplace kernel codebook	0.764	0.789	0.785	0.796	0.799	0.811	0.817	0.822	0.738
	Laplace codeword uncertainty	0.773	0.794	0.790	0.801	0.801	0.812	0.818	0.824	0.776
	Laplace codeword plausibility	0.724	0.759	0.740	0.726	0.725	0.737	0.718	0.731	0.696
L_2	Traditional codebook	0.704	0.785	0.748	0.768	0.757	0.734	0.724	0.727	0.743
	L_1-norm codebook	0.698	0.780	0.744	0.763	0.753	0.729	0.721	0.726	0.741
	L_2-norm codebook	0.687	0.733	0.717	0.739	0.744	0.729	0.722	0.724	0.729
	Gaussian kernel codebook	0.793	0.808	0.775	0.804	0.785	0.790	0.791	0.783	0.759
	Gaussian codeword uncertainty	0.755	0.799	0.762	0.773	0.767	0.772	0.764	0.765	0.792
	Gaussian codeword plausibility	0.705	0.805	0.764	0.776	0.771	0.735	0.722	0.727	0.746

where G_i is a gain value that depends on the relevance of the ith retrieved shape (1 if the shape that is the ith closest match to the query shape in the query shape's class, and 0 otherwise). Basically, DCG_i represents the relevance of the top-i results. Thus, the DCG score reflects the performance of the algorithm when correct results that are retrieved earlier are weighted higher than those retrieved later. All normalized DCG calculations are relative values in the interval [0, 1], and higher values imply better retrieval performance.

7.1.2 Descriptors

To compare the aforementioned spectral descriptors, we design a reasonable strategy to choose the best parameters for each descriptor in our experiments. As pointed out in [23], the overall performance of the dense descriptor computed for each point of the shape is superior than a sparse descriptor computed for a set of points produced by a feature detection algorithm. So, we compute all spectral descriptors on every vertex of the shape. In addition, due to the different ways of discretizing the LB operator, our implementation might generate slightly different results from the ones reported in [23]. However, for fair comparison, we use the same implementation for all the spectral methods in order to not favor any method.

7.1.3 Kernel size

We choose the best kernel size according to the predefined range underlying the data, from *Local Kernel Size* (σ_{Local})

Table 9 Performance (DCG) of different spectral descriptors on SHREC 2010 dataset using intrinsic spatial pyramid matching with various codebook models

Spectral descriptor	Clustering norm	Level L (Partitions)	Codebook models							
			Traditional		Kernel		Uncertainty		Plausibility	
			Single	Pyramid	Single	Pyramid	Single	Pyramid	Single	Pyramid
HKS	L_1	1 (2)	0.851	0.855	0.829	0.835	0.848	0.853	0.847	0.852
		2 (4)	0.856	0.863	0.843	0.849	0.858	0.862	0.855	0.860
		3 (8)	0.863	**0.867**	0.8554	**0.858**	0.865	0.867	0.861	**0.865**
		4 (16)	0.856	0.862	0.849	0.856	0.861	**0.867**	0.855	0.860
	L_2	1 (2)	0.837	0.842	0.845	0.849	0.845	0.849	0.847	0.842
		2 (4)	0.848	0.851	0.8616	0.866	0.859	0.860	0.850	0.853
		3 (8)	0.850	**0.853**	0.862	**0.866**	0.860	0.863	0.851	**0.855**
		4 (16)	0.847	0.852	0.854	0.862	0.859	**0.864**	0.847	0.852
SIHKS	L_1	1 (2)	0.799	0.798	0.7989	0.7992	0.8113	0.8107	0.7952	0.7951
		2 (4)	0.813	0.800	0.817	0.814	0.829	0.821	0.810	0.807
		3 (8)	0.818	0.810	0.823	0.821	0.829	0.828	0.815	0.812
		4 (16)	**0.831**	0.8262	**0.827**	0.825	**0.832**	0.8313	**0.830**	0.824
	L_2	1 (2)	0.872	0.873	0.853	0.854	0.874	0.876	0.873	0.874
		2 (4)	0.877	0.879	0.871	0.869	0.878	0.882	0.877	0.880
		3 (8)	0.879	0.881	0.872	0.874	0.878	0.8830	0.879	0.882
		4 (16)	0.883	**0.884**	0.876	**0.877**	0.884	**0.885**	0.883	**0.884**
HMS	L_1	1 (2)	0.748	0.747	0.722	0.725	0.756	0.759	0.739	0.743
		2 (4)	0.783	0.750	0.757	0.751	0.795	0.792	0.774	0.773
		3 (8)	0.787	0.781	0.778	0.778	0.802	**0.804**	0.780	**0.783**
		4 (16)	0.774	**0.790**	0.783	**0.786**	0.792	0.800	0.771	0.780
	L_2	1 (2)	0.749	0.752	0.737	0.740	0.776	0.780	0.749	0.752
		2 (4)	0.787	0.786	0.777	0.775	0.813	0.812	0.787	0.786
		3 (8)	0.793	**0.7971**	0.797	0.798	0.821	**0.824**	0.793	**0.797**
		4 (16)	0.783	0.792	0.795	**0.799**	0.811	0.820	0.783	0.792
WKS	L_1	1 (2)	0.725	0.726	0.711	0.711	0.725	0.727	0.724	0.725
		2 (4)	0.750	0.747	0.739	0.734	0.751	0.745	0.749	0.746
		3 (8)	0.759	0.757	0.748	0.745	**0.759**	0.756	0.758	0.757
		4 (16)	0.759	**0.760**	0.754	0.751	0.760	**0.761**	0.758	**0.760**
	L_2	1 (2)	0.728	0.729	0.671	0.670	0.672	0.670	0.728	0.729
		2 (4)	0.754	0.751	0.713	0.703	0.715	0.706	0.754	0.751
		3 (8)	0.762	0.761	0.713	0.721	0.731	0.725	0.762	0.761
		4 (16)	0.762	**0.763**	**0.732**	0.727	**0.735**	0.732	0.762	**0.763**
GPS	L_1	1 (2)	0.722	0.722	0.778	**0.789**	0.780	**0.790**	0.726	0.726
		2 (4)	0.717	0.720	0.735	0.747	0.735	0.745	0.717	0.721
		3 (8)	0.732	0.729	0.734	0.739	0.733	0.739	0.733	0.730
		4 (16)	**0.736**	0.734	0.726	0.731	0.726	0.731	0.737	**0.736**
	L_2	1 (2)	0.759	0.760	0.767	**0.778**	0.734	**0.745**	0.761	**0.772**
		2 (4)	0.759	0.762	0.723	0.744	0.722	0.729	0.728	0.753
		3 (8)	0.765	0.769	0.716	0.729	0.704	0.712	0.748	0.759
		4 (16)	0.768	**0.770**	0.720	0.725	0.691	0.697	0.751	0.758

Bold values indicate the best results

to *Global Kernel Size* (σ_{Global}). For each cluster, the distances from its centroid to other points of this cluster are computed. Then, the median absolute deviation is adopted to obtain the kernel size for each cluster. We use the median value as the local kernel size σ_{Local}. Since there is no explicit centroid for all the data points, we compute

Table 10 Performance comparison of descriptors and their optimal parameters on SHREC 2011 dataset

	Spectral descriptors						
	HKS	SIHKS	HMS	WKS	GPS	ShapeDNA	EVD
Parameters	$\tau = 1/4$	$\tau = 1/16$	$\tau = 2$	$M = 100$			
	$T = 5$	$T = 25$	$T = 40$	$\sigma = 0.05$			
	$t_0 = 0.01$	$F = 193$	$t_0 = 4$				
	$\alpha = 4$	$\alpha = 2$	$\alpha = 0.8$				
DCG	0.811	0.826	0.773	0.680	0.709	0.782	0.560

global "ghost centroid" in the same way that the centroid of each cluster is obtained during clustering. Then, the global kernel size σ_{Global} is estimated by treating the data space as a single cluster. σ_{Local} and σ_{Global} can help us to roughly estimate the distribution of the data. Recall that our goal is to find the best smoothing parameters in order to maximize the discrimination, not to fit the data best, we use them as the size markers of the data space, and get the scale $S = \sigma_{\text{Global}}/\sigma_{\text{Local}}$. Multiplying by an augmenting coefficient A, we set an interval $[\sigma_{\text{Local}}/(AS), \sigma_{\text{Local}}AS]$ as the candidate space. With a sampling step $F_A \propto A$ to discretize the space, we select the best $\hat{\sigma}$ as the final kernel size. Note that in order to be consistent with the clustering stage, when estimating the kernel size, we use the L_1-norm for the dictionary learned by the cityblock distance, and each centroid is the component-wise median of the points in that cluster. We use the L_2-norm for the one learned by the Euclidean distance, and each centroid is computed as the mean of the points in that cluster. In Table 2, the parameters of different spectral descriptors are set for each dataset to get the interval of the kernel size.

7.1.4 Complexity

The codes were implemented in MATLAB 7.14 (R2012a). The experiments were performed on a desktop computer with an Intel Core i3-2100 running at 3.1 GHz and 4 GB RAM. The computation of the vocabulary is performed offline in advance. It depends on the number of the descriptors (number of vertices), the dimension of the descriptor, and the vocabulary size (the number of clusters). Since we simplify our mesh to 2000 faces for each shape, we have a set of approximately 2×10^5 descriptors. Since the 3D models used in our experiments are watertight, simplifying a surface mesh by reducing its number of faces to 2000 would essentially preserve the shape semantically. Consequently, the discriminative power of the descriptors would not change drastically after mesh simplification.

To confirm getting optimal results, the clustering is repeated 3 times, and each by a new set of initial cluster centroid positions. The solution with the lowest value for the sum of distances is returned. In Table 3, we list the runtime in seconds for various descriptor dimensions and vocabulary sizes.

7.2 Results on SHREC 2010 dataset

7.2.1 In-depth analysis of descriptors and ambiguity

The first dataset we consider is SHREC 2010 [70], which is a standard dataset of nonrigid 3D models used in the Shape Retrieval Contest, organized by National Institute of the Standards and Technology (NIST). The dataset consists of 200 shapes spread over 10 categories with 20 shapes each, and range from human body to man-made tools like glasses. Some of the deformations performed on the shapes are artificially generated, which would result in misleading recognition. In Fig. 6, we show 4 models of each class in this dataset.

For the SHREC 2010 dataset, we analyze the types of spectral descriptors, vocabulary size, and codeword ambiguity. The vocabulary sizes we consider are 8, 12, 16, 24, 32, 48, 64, and 80. To gain further insight into the performance variation between the various types of spectral descriptors, we show the retrieval results for different parameters in Tables 13, 14, 15, 16 (see "Appendix"). The L_2-norm traditional codebook is used in all the descriptors. The main goal of our parameter determination is to select the appropriate heat diffusion time for each descriptor in order to maximally discriminate between the shapes. For HKS, we formulate the diffusion time as $t = t_0\alpha^{\tau}$, where τ is sampled from 0 to a given *scale* with a resolution 1/4 in our case. We highlight the best result for each factor t_0, and notice that the largest diffusion times are $t_{\max} = t_0\alpha^{\max\tau} = \{343.44, 131.84, 74.08, 238.15, 139.57\}$. These times are obviously different from the best parameters $t = \{1{,}024, 1{,}351, 1{,}783, 2{,}353, 3{,}104, 4{,}096\}$ used in [23]. Since we use a different dataset, the difference of best parameters is reasonable. If the confusion time is larger than $t_{\max}$, i.e., the heat diffuses for enough long time, then the heat distribution of the whole shape will be very

Table 11 Performance (DCG) of different spectral descriptors on SHREC 2011 dataset using intrinsic spatial pyramid matching with varying codebook models

Spectral descriptor	Clustering norm	Level L (Partitions)	Codebook models							
			Traditional		Kernel		Uncertainty		Plausibility	
			Single	Pyramid	Single	Pyramid	Single	Pyramid	Single	Pyramid
HKS	L_1	1 (2)	0.829	0.822	0.811	0.806	0.831	0.824	0.808	0.802
		4 (16)	0.878	0.872	0.871	0.866	0.883	0.878	0.874	0.867
		7 (128)	0.890	0.888	0.880	0.879	0.892	0.891	0.887	0.885
		9 (512)	**0.893**	0.892	0.879	**0.880**	**0.894**	0.893	**0.890**	0.889
	L_2	1 (2)	0.834	0.827	0.812	0.807	0.849	0.843	0.834	0.827
		4 (16)	0.876	0.872	0.879	0.875	0.889	0.887	0.876	0.872
		7 (128)	0.889	0.888	0.878	**0.880**	0.889	**0.890**	0.889	0.887
		9 (512)	**0.891**	0.890	0.868	0.873	0.888	0.889	**0.891**	0.890
SIHKS	L_1	1 (2)	0.853	0.849	0.873	0.870	0.861	0.857	0.852	0.850
		4 (16)	0.878	0.876	0.888	0.888	0.887	0.884	0.878	0.876
		7 (128)	0.889	0.888	0.895	0.894	0.890	0.895	0.889	0.887
		9 (512)	**0.891**	0.890	**0.896**	0.896	**0.898**	0.897	**0.891**	0.890
	L_2	1 (2)	0.845	0.843	0.868	0.864	0.871	0.866	0.845	0.843
		4 (16)	0.869	0.867	0.885	0.885	0.886	0.885	0.869	0.867
		7 (128)	0.878	0.877	**0.889**	0.888	0.888	**0.889**	0.878	0.877
		9 (512)	**0.880**	0.879	0.888	0.888	0.889	0.889	**0.880**	0.879
HMS	L_1	1 (2)	0.780	0.784	0.759	0.764	0.801	0.804	0.777	0.781
		4 (16)	0.823	**0.828**	0.833	**0.837**	0.845	**0.850**	0.821	**0.826**
		7 (128)	0.815	0.819	0.819	0.814	0.831	0.835	0.814	0.818
		9 (512)	0.825	0.823	0.804	0.803	0.836	0.836	0.825	0.822
	L_2	1 (2)	0.782	0.786	0.777	0.779	0.794	0.798	0.782	0.786
		4 (16)	0.827	**0.837**	0.831	**0.838**	0.835	**0.842**	0.827	**0.832**
		7 (128)	0.821	0.824	0.806	0.814	0.823	0.828	0.821	0.825
		9 (512)	0.830	0.828	0.796	0.803	0.828	0.828	0.830	0.828
WKS	L_1	1 (2)	0.757	0.746	0.777	0.738	0.765	0.755	0.751	0.741
		4 (16)	0.829	0.824	**0.834**	0.828	**0.841**	0.838	0.828	0.723
		7 (128)	0.839	0.8382	0.824	0.828	0.826	0.829	0.838	0.837
		9 (512)	0.839	**0.839**	0.813	0.818	0.819	0.822	0.839	**0.839**
	L_2	1 (2)	0.718	0.709	0.761	0.754	0.781	0.776	0.678	0.707
		4 (16)	0.797	0.793	0.813	**0.823**	0.822	**0.830**	0.797	0.792
		7 (128)	0.804	**0.804**	0.779	0.789	0.787	0.796	0.803	**0.804**
		9 (512)	0.800	0.803	0.756	0.764	0.776	0.782	0.800	0.802
GPS	L1	1 (2)	0.691	0.693	0.763	**0.777**	0.764	**0.776**	0.691	0.693
		4 (16)	0.712	0.711	0.692	0.700	0.691	0.670	0.712	0.712
		7 (128)	0.718	0.718	0.684	0.686	0.683	0.686	0.719	0.718
		9 (512)	**0.719**	0.719	0.681	0.683	0.680	0.682	**0.720**	0.719
	L_2	1 (2)	0.705	0.708	0.752	**0.759**	0.751	**0.757**	0.705	0.708
		4 (16)	0.726	0.725	0.705	0.716	0.705	0.715	0.726	0.725
		7 (128)	0.728	0.729	0.678	0.683	0.678	0.683	0.729	0.729
		9 (512)	**0.729**	0.729	0.671	0.675	0.670	0.674	**0.729**	0.729

Bold values indicate the best results

similar, which tends to degrade the discrimination. On the contrary, if the confusion time is smaller than $t_{\max}$, then only local patches of the shape are considered in the description, thus the global description of the whole shape is deficient. In order to construct the SIHKS, we use $t = \alpha^{\tau}$, where τ ranges from 1 to a given scale with finer

Table 12 Summary comparison of spectral signatures

Property	SIHKS	HKS	WKS	HMS	GPS
Discriminative power	1	2	5	3	4
Compactness	5	1	2	3	4
Localization	4	3	1	2	5
Ambiguity	3	5	2	4	1

increments of 1/16. After applying the logarithm, derivative, and Fourier transform, the first several discrete lowest frequencies are used as the local descriptor. In [23], the first 6 lowest frequencies are adopted, which yield a satisfactory result on SHREC 2010 dataset as shown in Table 14. To guarantee not to favor any method in our experiments, we test with various frequencies and find that the best result

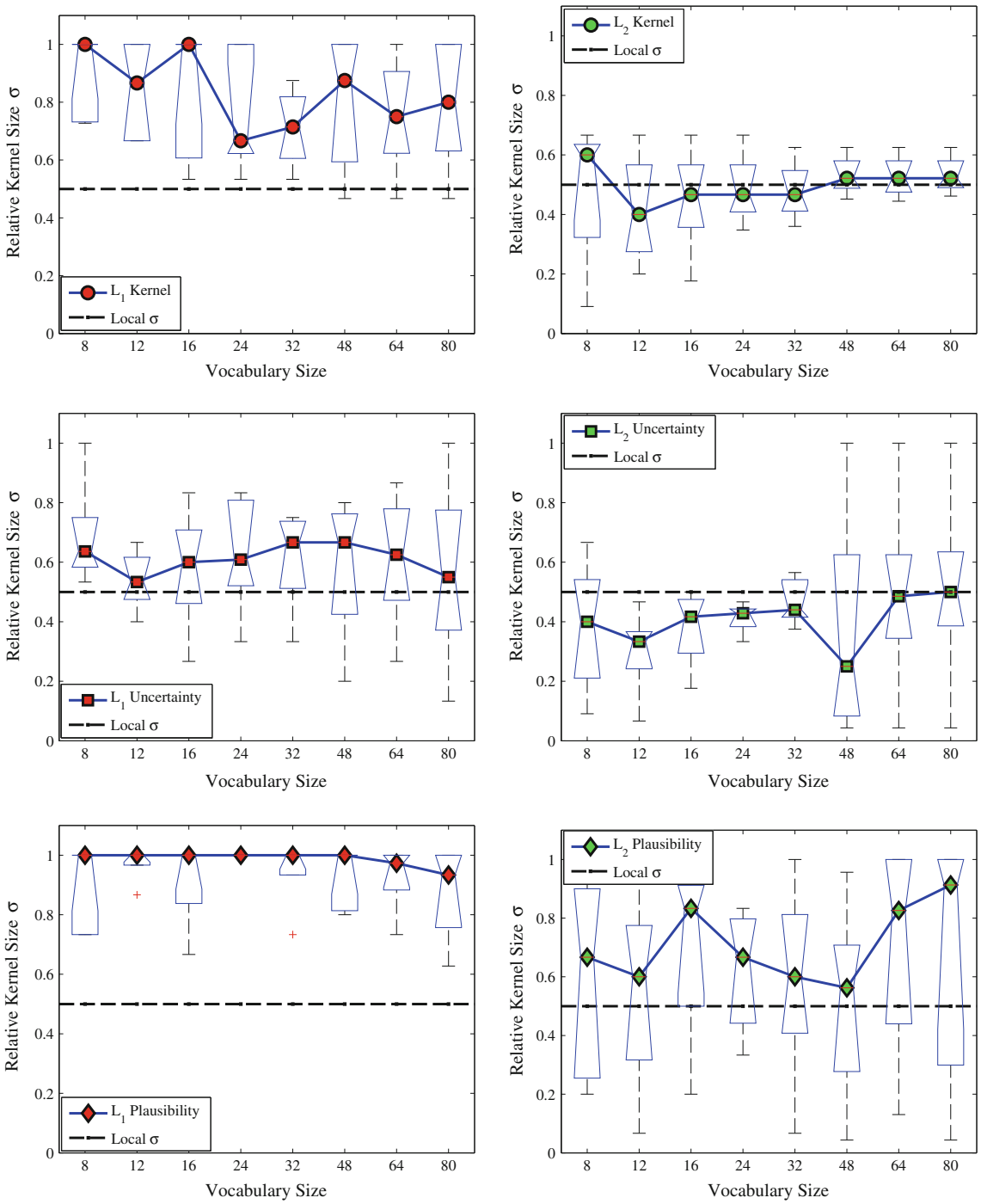


Fig. 7 Illustration of relative kernel size of different ambiguity modeling methods. *Top* kernel codebook. *Middle* codeword uncertainty. *Bottom* codeword plausibility

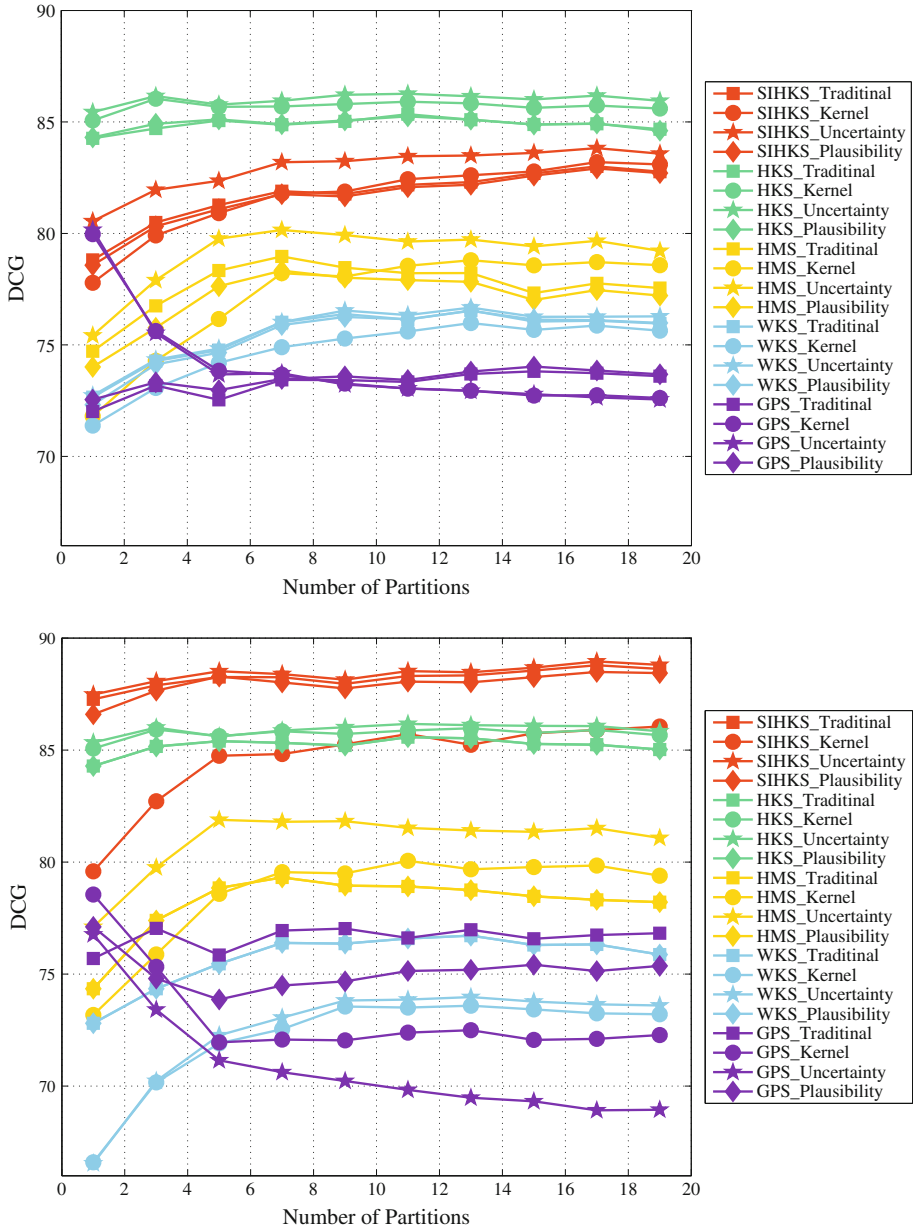
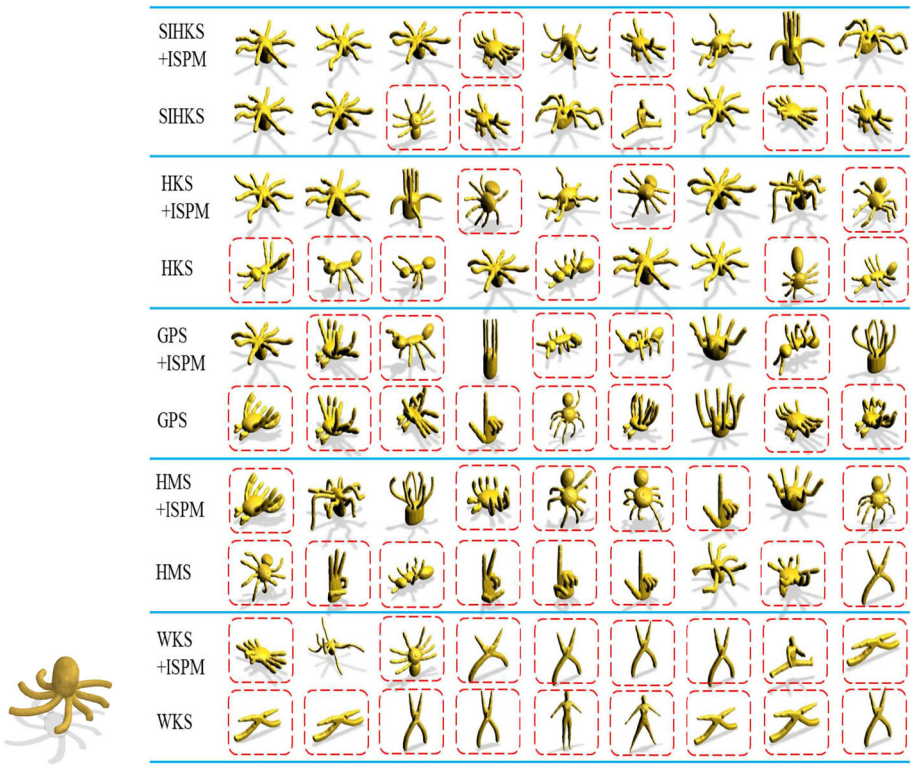


Fig. 8 Performance improvement by increasing the number of intrinsic partitions on SHREC 2010 dataset. *Top* L_1-norm. *Bottom* L_2-norm

can always be obtained when all the frequencies are used. One-dimensional HMS is used to construct the temperature distribution descriptor [17]. For mesh segmentation, HMS prefers small time because high-resolution details are preserved at a small scale. Our goal is to discriminate between shapes, so we construct a multi-scale HMS to compare the temperature distribution with multiple diffusion times. By defining a universal time unit δ as in the computation of the TD descriptor, we use $t = \tau\delta$ with τ ranging from 1 to a given scale to compute the descriptor. In Table 15, $\delta = 1$ and $t = [\delta, 2\delta, \ldots, 100\delta]$ give the best result for the multi-scale HMS. The WKS is associated with each vertex on the surface, and it describes the energy distributions over a family of fixed energy scales. There are two parameters in this descriptor, namely the increment δ and the variance σ. Since the energy range $[\lambda_{\min}, \lambda_{\max}]$ is determined by the eigenvalues and it is fixed, we seek the best increment by dividing the interval by a scalar M. The variance σ is also inferred as certain percentage of the interval. As shown in Table 16, the best parameters are $M = 50$ and percentage equal to 0.2. Finally, the best parameters for each spectral descriptor are summarized in Table 17 in the Appendix. SIHKS performs the best, and is slightly better than HKS since the variance of the scale transformation is small in SHREC 2010. As expected, the worst performance is obtained by WKS, which characterizes only local

(a) Query shape spectacle and its retrieval results



(b) Query shape octopus and its retrieval results

Fig. 9 Retrieval results using different spectral descriptors and its spatial enhanced version. Error result is marked in the *red dashed box*. **a** In the *left* is the query-shaped spectacle, and the ten rows in the *right* its *top 9* retrieval results; **b** In the *left* is the query-shaped octopus, and the ten rows in the *right* its *top 9* retrieval results

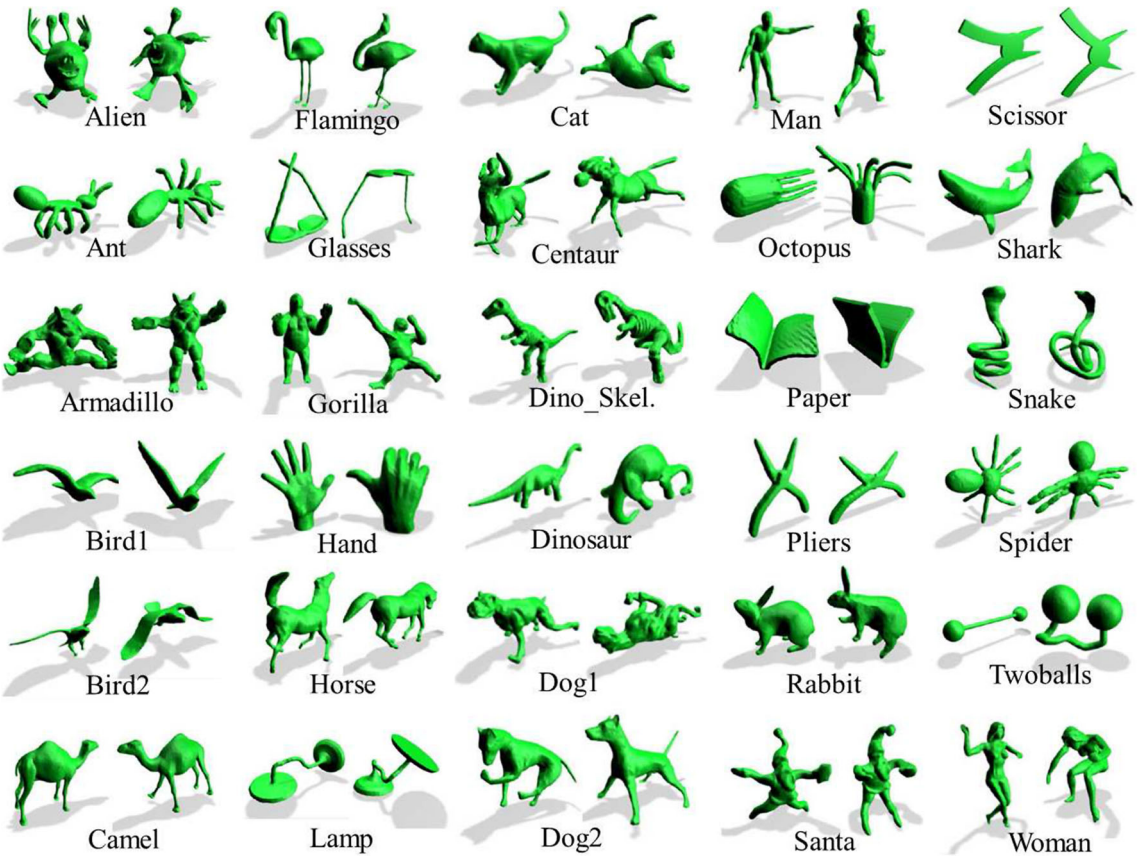


Fig. 10 Sample shapes in SHREC 2011 dataset

information of the small patch. We also considered a 200 vocabulary size for each signature, and the results remain virtually unchanged or go down slightly compared to a smaller vocabulary size.

In Table 4, 5, 6, 7 we list the DCG values for different spectral signatures and classes on SHREC 2010. It should be noted that all of the signatures are good at retrieving the Teddy model. SIHKS yields the highest accuracy on most of the classes, such as ant, crab, hand, human, octopus, plier, spectacle, and spider. HKS is slightly lower than SIHKS. Surprisingly, HMS is the best one in retrieving the Teddy shape. Also, the comparison between WKS, HMS, and GPS is inspiring. WKS performs well on octopus, snake, and spectacle. These classes have a high percentage of thin-branch-like components. GPS gets a relatively high accuracy on ant, hand, and human.

To study the influence of the vocabulary size, we use various local descriptors and change the vocabulary size from 8 to 80 geometric words. Tables 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 show the resulting performance. As can be observed, the overall performance improves with the increase of the vocabulary size, but at the expense of the representation size (length of the BoF vector). However, for HKS and GPS, the best performance is obtained when the vocabulary size is 32 and 12, respectively.

Tables 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 also show the retrieval results with the various types of codeword ambiguity. The results show that codeword uncertainty consistently outperforms other types of ambiguity for all kinds of descriptors and all vocabulary sizes. Besides the 4 types of ambiguity modeling methods in [44], we also include the accumulation of L_1 and L_2 norms from descriptor to codeword. The reason is twofold: (1) to keep coherent with the clustering stage, and (2) VLAD, which is the accumulation of vector from descriptor to codeword, achieves better results as reported in the literature. Indeed, in our experiment, SIHKS, HMS, and WKS with distance accumulation outperform the traditional codebook, and even are superior than the codebook uncertainty in certain cases. In addition, the L_1- and L_2-norms in the clustering stage have a considerable effect on the performance. For GPS, the L_1-norm based methods are much worse than L_2-norm based ones. But for WKS, the L_1-norm based method is better than L_2-norm based one. As a result, we may conclude that the WKS descriptor is more discriminative than the L_1-norm.

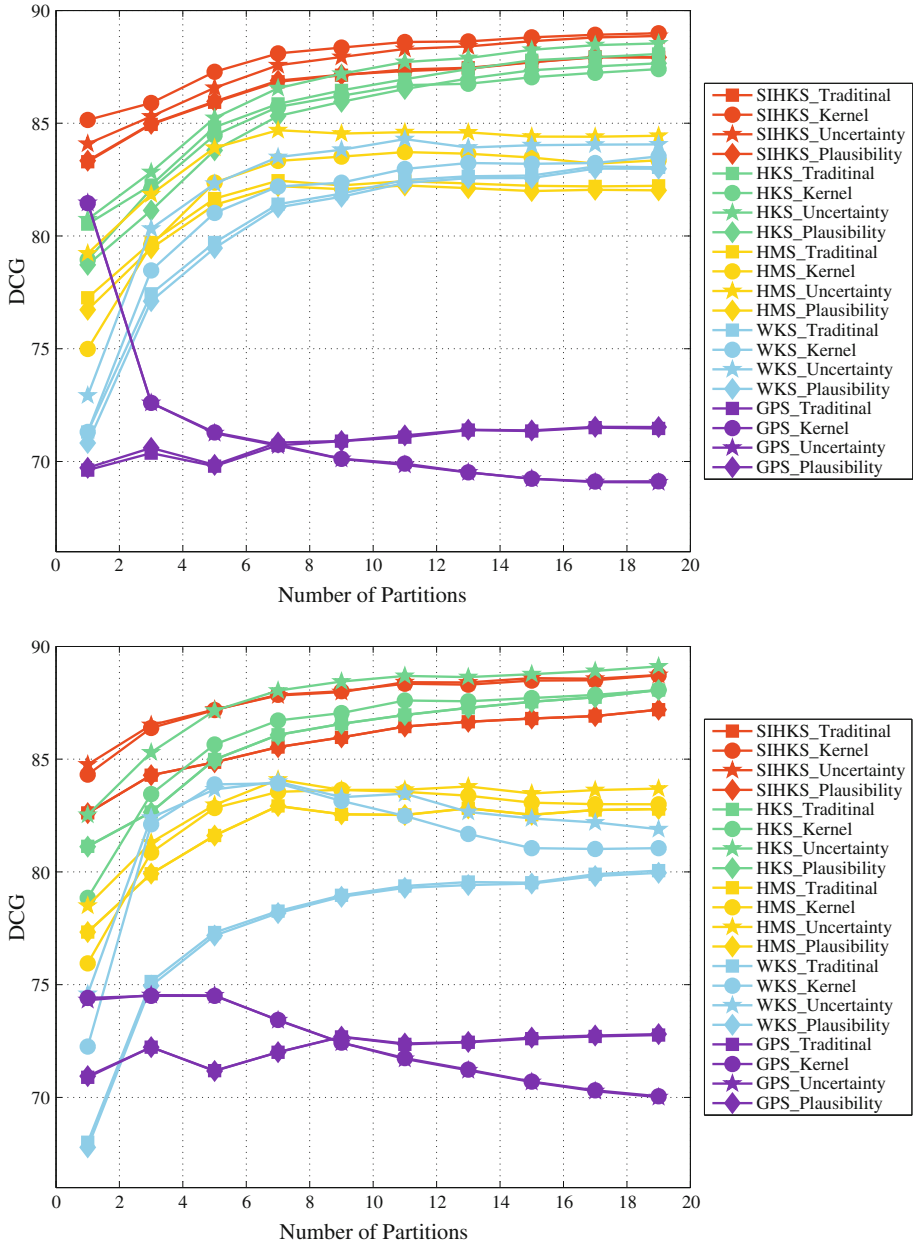


Fig. 11 Performance improvement by increasing the number of intrinsic partitions on SHREC 2011 dataset. *Top* L_1-norm. *Bottom* L_2-norm

To examine the influence of the kernel size, we show the kernel size found as the statistic among the five spectral descriptors. Following the technique of optimal kernel size estimation described in the experimental setting (Subsect. 7.1), σ_{Local} is calculated first. Then, the optimal kernel size $\hat{\sigma}$ is determined experimentally inside the interval for each vocabulary size. To better visualize the results, we obtain the relative $\hat{\sigma}$ as the relative position in the interval. Thus, the relative σ_{Local} is 0.5. Our goal is to compare $\hat{\sigma}$ for different ambiguity types in both Laplacian and Gaussian kernels. So, the five number summary is computed on the set composed of the relative $\hat{\sigma}$ of five descriptors in the same vocabulary setting, and the results are shown in Fig. 7. For both Laplacian and Gaussian kernels, the codeword plausibility has the largest kernel size, followed by the kernel codebook; the smallest kernel size is held by the codeword uncertainty. The Laplacian kernel has a larger kernel size than the Gaussian kernel for each type of ambiguity methods. In our experiment, we also found that increasing the kernel size of the codeword plausibility beyond a sufficiently large value does not significantly change the accuracy. The kernel size of the Laplacian codeword plausibility remains the largest in the interval, but the kernel size of the Gaussian codeword plausibility among different descriptors oscillates the most. Note that since we plot the relative position of $\hat{\sigma}$ compared to

σ_{Local}, the value of σ_{Local} becomes smaller over the number of vocabulary elements. In fact, it shows that a larger vocabulary leads to slightly smaller kernels. This phenomenon is consistent with the image domain [44]. This may be expected since a larger vocabulary is formed by a smaller radius between codewords.

7.2.2 Improvement with intrinsic partition

First, we examine the effect of integrating spatial cues on surfaces via the intrinsic partition. Figure 8 shows the performance improvement of the retrieval experiments by matching shapes directly using intrinsic partitions on SHREC 2010 dataset. With the increase of the number of intrinsic partitions, all the spectral descriptors, except GPS, are improved substantially in a global fashion, even though the performance drops down in certain numbers. For ShapeGoogle [23], its performance is plotted as points whose partition number is one. Obviously, intrinsic spatial cues on shape surface proposed in our framework significantly outperform ShapeGoogle. We conjecture that GPS is degraded because of its global nature. A crucial parameter is the number of partitions. Experimentally, we find that the accuracy remains stable after 16 partitions for all the spectral descriptors.

Next, let us examine the behavior of intrinsic spatial pyramid matching. For completeness, Table 9 lists the performance achieved using just the highest level of the pyramid (the "single" columns) as well as the performance of the complete matching scheme using multiple levels (the "pyramid" columns). For all kinds of features, except GPS, the results improve considerably as we go from $L = 1$ to a multi-level setup. We do not display the results for $L = 0$ because its highest single level is the same as with its pyramid. Although matching at the highest pyramid level seems to account for most of the improvement, using all the levels together yields a statistically significant benefit. For strong features, single-level performance actually drops as we go from $L = 3$ to $L = 4$. This means that the highest level of the $L = 3$ pyramid is too finely subdivided, with individual bins yielding few matches. Despite the diminished discriminative power of the highest level, the performance of the entire $L = 4$ pyramid remains essentially identical to that of the $L = 3$ pyramid. Thus, the main advantage of the intrinsic spatial pyramid representation stems from combining multiple resolutions in a principled fashion, and it is robust to failures at individual levels Tables 6, 7.

It is also important to compare the performance of different spectral descriptor sets. As expected, weak descriptors do not perform as well as strong descriptors, though in combination with the spatial pyramid, they can also achieve acceptable levels of accuracy. Note that only descriptors with a much higher density and much smaller spatial extent will continue to improve their performance as we increase L from 3 to 4. Such kinds of descriptors in Table 9 include SIHKS and WKS. In this respect, intrinsic spatial pyramid matching (ISPM) provides us with a way to analyze the spectral descriptors. On the other hand, the performances of HKS and HMS drop when the pyramid level increases from 3 to 4. Moreover, the performance of GPS decreases immediately when the descriptors are aggregated in terms of local patches instead of the whole shape. Increasing the visual vocabulary size from 8 to 80 might result in a small performance increase at $L = 0$, but this difference is eliminated at higher pyramid levels. Thus, we may conclude that the coarse-grained geometric cues provided by the pyramid have more discriminative power than an enlarged geometric vocabulary. Another explanation for the improvement is that the geometric cues eliminate the word ambiguity in a spatial context, and its ambiguity modeling ability is also more discriminative than the ambiguity codebook models. For example, HMS with L_2-norm, the traditional codeword, achieves 0.743 on SHREC 2010 dataset, and the improvement brought by the codeword uncertainty is $0.771 - 0.743 = 0.028$. However, the improvement brought by ISPM is $0.792 - 0.743 = 0.049$.

In Fig. 9, we show two examples of top 9 retrieval results for different methods. There are plenty of examples to demonstrate that our proposed ISPM method improves the performance of the original codebook models with varied spectral descriptors. However, in order to illustrate the merits of different descriptors, we choose the two exemplar queries that bring consistent results with the conclusion we get for the spectral descriptors. For both queries, the ISPM method enhances the results by helping rank more relevant shapes higher as expected, and SIHKS combined with ISPM achieves a more satisfactory result. For the first query spectacle, the primitive descriptors always confuse it with lines, pliers, human body, and octopus. This is because these objects also have several long, thin pipe-like parts and flat globular parts, and the proportions are similar. The spatial partition integrates the intrinsic symmetry cue, and separates pipe-like parts and globular parts into different sub-histograms, thus resulting in a more descriptive representation. The pipe-like parts attached to fringe of shapes are almost similar among these classes, but the globular parts in the middle have different degree and amount of flat, which is the key to distinguish the spectacles. For the second query octopus, multiple legs objects like spiders and ants are usually retrieved mistakenly by the descriptors SIHKS, HKS, and GPS. This is in fact understandable since even humans would consider

these shapes as similar from their appearance. HMS ranks some hands higher, and WKS ranks some pliers higher. This phenomenon shows that the different spectral descriptors characterize different aspects of shape features.

It is worth pointing out that the intrinsic spatial partition helps measure the localization ability of the signatures, meaning the ability of a signature to capture the local geometry around the mesh vertex. The DCG improvement after applying spatial partition reflects localization due largely to the fact that the partition separates the signature into different patches. The finer is the partition, the more details can be captured if the DCG improves.

7.3 Results on SHREC 2011

Our second set of experiments is performed on a larger dataset, SHREC 2011 [71], which contains 600 watertight triangle meshes that are equally classified into 30 categories. SHREC 2011 is the most diverse nonrigid 3D shape database available today in terms of object classes and deformations. In Fig. 10, we show 2 models of each class in the dataset.

We follow the same experimental procedure on SHREC 2010 dataset. First, we find the best performance of each spectral descriptor with its best parameters under the traditional codebook model, as shown in Table 10. Then, using these parameters we conduct further experiments on various ambiguity modeling models. In this way, we get the baseline result, which is actually the algorithm of ShapeGoogle [23]. Finally, our spatially aggregating approach is tested against each type of descriptors. In Fig. 11, we display the performance improvement with the increasing number of partitions. For both L_1- and L_2-norms, all types of descriptors, except GPS, show improved accuracy when spatially aggregated. HKS and SIHKS are improved throughout all the number of partitions till 512 in our experiment. But we only show the results from 1 to 20 for the sake of visualization. HMS and WKS remain unchanged or become worse when 10 or more partitions are adopted. Note that an interesting phenomenon is that GPS with traditional and plausibility ambiguity has a lower accuracy than kernel and uncertainty ambiguity when no global spatial is integrated. However, the results are completely the opposite after a certain degree of partition is included. The traditional and plausibility ambiguity modeling methods go up, but kernel and uncertainty ambiguity modeling methods go down, even lower than the former approaches.

Table 11 gives a breakdown of retrieval accuracies for different pyramid levels and different spectral descriptors with vocabulary size 32. On this large dataset, a major advantage of ISPM is shown to provide a robust and stable performance by weightily combining multiple spatial levels, especially, for HMS and WKS, and it consistently outperforms the single level approach. Although ISPM is not superior in all cases, it is, however, practical since determining the level of partition is much easier than determining the number of partitions to obtain satisfactory results.

In summary, our proposed shape retrieval approach is shown to outperform state-of-the-art orderless descriptor aggregating methods. Because of the geometric stability and lack of damage of shapes in SHREC 2011, dense descriptors combined with global spatial relations seem to capture more discriminative information about the objects by providing an approximate correspondence.

To gain further insight into the strengths and weaknesses of each descriptor, we compared the shape signatures in terms of four different properties: discriminative power, compactness, localization, and ambiguity. Discriminative power describes the signature's ability to distinguish between shapes belonging to different classes. Since we used watertight meshes from different classes in our experiments, the DCG value reflects the discriminative power of the descriptors. Compactness refers to the dimension that a signature has to achieve for its maximum discriminative power. A lower dimension indicates higher compactness, and leads to simpler computation. We introduce localization as the ability of a signature to capture the local geometry around the mesh vertex. The DCG improvement after spatial partition reflects localization, since the partition separates the signatures into different patches. The finer is the partition, the more details can be captured if the DCG value further improves. To describe the distribution of the signature in the feature space, we observe the clustering centroid of the signature. The higher is the ambiguity between geometric words, the more kernel uncertainty is beyond the L_0-norm codebook. In Table 12, we rank the shape signatures in terms of different properties on a scale from 1 to 5. A value of 1 means the strongest, while a value of 5 means the weakest.

8 Conclusion and future work

This paper reviewed and compared five recent spectral descriptors and varied codebook ambiguity models for nonrigid 3D shape retrieval. It turns out that SIHKS is the most discriminative spectral descriptor, and that codeword uncertainty yields the best ambiguity modeling ability between codewords, without taking codeword plausibility into account. By integrating the spatial cues with the proposed intrinsic partition, the retrieval performance was significantly improved. The intrinsic spatial pyramid

matching, which works by repeatedly partitioning the surface and computing histograms of spectral descriptors over the resulting subpatches, showed promising results on two standard datasets of 3D models.

We believe that the optimal way for exploiting ambiguity modeling both in the shape and in the feature space in a unified framework would result in higher performance; this is a subject that we intend to explore as future work.

Appendix

See Tables 13, 14, 15, 16, 17 and 18.

Table 13 The DCG measure of HKS based on different pairs of parameters

HKS		α				
Factor t_0	Scale	1.24	1.64	2	2.5	3
4.24	2	0.701	0.741	0.765	0.791	0.803
	4	0.723	0.778	0.802	0.830	**0.831**
	6	0.746	0.801	0.817	0.827	0.822
	8	0.756	0.805	0.823	0.822	0.817
	10	0.765	0.811	0.820	0.811	0.820
	12	0.775	0.805	0.811	0.812	0.815
8.24	2	0.752	0.781	0.812	0.835	0.827
	4	0.781	0.821	**0.840**	0.839	0.829
	6	0.780	0.831	0.837	0.827	0.823
	8	0.810	0.833	0.828	0.825	0.814
	10	0.814	0.832	0.831	0.819	0.817
	12	0.824	0.829	0.817	0.817	0.811
10.24	2	0.763	0.799	0.817	0.835	0.843
	4	0.793	**0.844**	0.833	0.838	0.829
	6	0.806	0.843	0.836	0.834	0.824
	8	0.823	0.840	0.837	0.825	0.815
	10	0.826	0.828	0.826	0.813	0.814
	12	0.836	0.835	0.816	0.816	0.810
12.24	2	0.772	0.807	0.821	0.830	0.829
	4	0.791	0.840	0.831	0.829	0.823
	6	0.817	**0.846**	0.832	0.822	0.815
	8	0.833	0.834	0.827	0.820	0.815
	10	0.842	0.835	0.830	0.814	0.819
	12	0.839	0.828	0.820	0.811	0.808
16.24	2	0.787	0.810	0.809	0.822	0.832
	4	0.805	0.828	0.835	0.819	0.821
	6	0.827	0.836	0.819	0.813	0.814
	8	0.844	0.825	0.819	0.810	0.803
	10	**0.848**	0.824	0.822	0.810	0.813
	12	0.835	0.820	0.813	0.811	0.799

Bold values indicate the best results

Table 14 The DCG measure of SIHKS based on different pairs of parameters

SIHKS		α				
Timescale	Frequency	1.24	1.64	2	2.5	3
20	6	0.795	0.842	0.867	0.853	0.846
	50	0.792	0.851	0.861	0.847	0.847
	100	0.794	0.847	0.869	0.852	0.845
	150	0.793	0.852	0.864	0.853	0.854
	193	0.795	0.852	0.866	0.848	0.855
25	6	0.774	0.867	0.872	0.854	0.849
	50	0.770	0.859	0.869	0.864	0.852
	100	0.769	0.861	0.871	0.862	0.851
	150	0.764	0.861	0.862	0.864	0.848
	193	0.757	0.861	**0.877**	0.862	0.845
30	6	0.811	0.860	0.859	0.857	0.850
	50	0.808	0.869	0.868	0.857	0.853
	100	0.819	0.872	0.865	0.863	0.856
	150	0.809	0.866	0.861	0.859	0.854
	193	0.809	0.868	0.870	0.853	0.854

Bold value indicates the best result

Table 15 The DCG measure of HMS based on different pairs of parameters

HMS	τ							
τ	40τ	60τ	80τ	100τ	120τ	140τ	160τ	180τ
0.005	0.676	0.692	0.697	0.690	0.734	0.712	0.729	0.740
0.01	0.702	0.722	0.744	0.714	0.708	0.734	0.712	0.714
0.02	0.713	0.726	0.712	0.742	0.719	0.720	0.722	0.723
0.05	0.724	0.734	0.729	0.744	0.736	0.733	0.727	0.737
0.1	0.744	0.739	0.742	0.717	0.712	0.729	0.713	0.705
0.2	0.723	0.716	0.712	0.718	0.749	0.753	0.751	0.753
0.4	0.739	0.744	0.744	0.753	0.749	0.753	0.745	0.752
0.6	0.735	0.747	0.752	0.750	0.747	0.754	0.747	0.742
1	0.751	0.751	0.753	**0.754**	0.744	0.744	0.739	0.738
2	0.739	0.744	0.729	0.733	0.723	0.736	0.731	0.725

Bold value indicates the best result

Table 16 The DCG measure of WKS based on different pairs of parameters

WKS	σ				
M	0.12	0.16	0.20	0.24	0.28
20	0.708	0.720	0.718	0.724	0.697
50	0.722	0.724	**0.727**	0.725	0.703
80	0.713	0.720	0.725	0.724	0.706
100	0.717	0.712	0.726	0.720	0.708
120	0.720	0.720	0.718	0.715	0.713
150	0.716	0.724	0.723	0.713	0.714

Bold value indicates the best result

Table 17 Performance comparison of descriptors and their optimal parameters on SHREC 2010 dataset

	Spectral descriptors						
	HKS	SIHKS	HMS	WKS	GPS	ShapeDNA	EVD
Parameters	$\tau = 1/4$	$\tau = 1/16$	$\tau = 2$	$M = 50$			
	$T = 10$	$T = 25$	$t = 100\tau$	$\sigma = 0.2$			
	$t_0 = 16.24$	$F = 193$					
	$\alpha = 1.24$	$\alpha = 2$					
DCG	0.848	0.877	0.754	0.727	0.757	0.801	0.636

Table 18 Performance (DCG) using different codebook models of varying size based on HKS local descriptor

Codebook model		Vocabulary size								
Clustering	Coding	8	12	16	24	32	48	64	80	200
L_1	Traditional codebook	0.827	0.841	0.852	0.846	0.854	0.836	0.828	0.837	0.834
	L_1-norm codebook	0.812	0.825	0.839	0.808	0.848	0.794	0.776	0.823	0.813
	L_2-norm codebook	0.784	0.793	0.812	0.765	0.808	0.746	0.726	0.774	0.755
	Laplace kernel codebook	0.820	0.832	0.840	0.852	0.841	0.851	0.847	0.845	0.831
	Laplace codeword uncertainty	0.829	0.843	0.853	0.849	0.857	0.840	0.839	0.848	0.834
	Laplace codeword plausibility	0.829	0.840	0.849	0.851	0.852	0.844	0.843	0.836	0.834
L_2	Traditional codebook	0.826	0.851	0.843	0.845	0.845	0.834	0.831	0.828	0.834
	L_1-norm codebook	0.804	0.838	0.823	0.837	0.825	0.791	0.806	0.785	0.820
	L_2-norm codebook	0.776	0.807	0.788	0.813	0.794	0.748	0.770	0.741	0.789
	Gaussian kernel codebook	0.826	0.846	0.847	0.845	0.851	0.852	0.848	0.849	0.829
	Gaussian codeword uncertainty	0.828	0.845	0.845	0.852	0.851	0.836	0.834	0.833	0.829
	Gaussian codeword plausibility	0.831	0.852	0.846	0.845	0.846	0.842	0.837	0.840	0.845

References

1. Yang, Y., Lin, H., Zhang, Y.: Content-based 3-D model retrieval: a survey. IEEE Trans. Syst. Man Cybern. Part C **37**(6), 1081–1098 (2007)
2. Del Bimbo, A., Pala, P.: Content-based retrieval of 3D models. ACM Trans. Multimedia Comput. Commun. Appl. **2**(1), 20–43 (2006)
3. Tangelder, J.W.H., Veltkamp, R.C.: A survey of content based 3D shape retrieval methods. Multimedia Tools Appl. **39**(3), 441–471 (2008)
4. Bustos, B., Keim, D.A., Saupe, D., Schreck, T., Vranic, D.V.: Feature-based similarity search in 3D object databases. ACM Comput. Surv. **37**(4), 345–387 (2005)
5. Jain, V., Zhang, H.: A spectral approach to shape-based retrieval of articulated 3D models. Comput. Aided Design **39**(5), 398–407 (2007)
6. Macrini, D., Siddiqi, K., Dickinson, S.J.: From skeletons to bone graphs: medial abstraction for object recognition. In: Proc. CVPR, pp. 1–8 (2008)
7. Siddiqi, K., Zhang, J., Macrini, D., Shokoufandeh, A., Bouix, S., Dickinson, S.J.: Retrieving articulated 3-D models using medial surfaces. Mach. Vis. Appl. **19**(4), 261−275 (2008)
8. Siddiqi, K., Pizer, S. (eds.): Medial representations: mathematics, algorithms and applications. Springer, Berlin (2008)
9. Li, C., Ben Hamza, A.: Skeleton path based approach for nonrigid 3D shape analysis and retrieval. In: Proc. IWCIA, LNCS, pp. 84–95 (2011)
10. Mohamed, W., Ben Hamza, A.: Reeb graph path dissimilarity for 3D object matching and retrieval. V. Comput. **28**(3), 305–318 (2012)
11. Sun, J., Ovsjanikov, M., Guibas, L.J.: A concise and provably informative multi-scale signature based on heat diffusion. Comput. Graph. Forum **28**(5), 1383–1392 (2009)
12. Kokkinos, I., Bronstein, M.M., Yuille, A.: Dense scale-invariant descriptors for images and surfaces. Research Report, INRIA RR-7914 (2012)
13. Fang, Y., Sun, M., Kim, M., Ramani, K.: Heat-mapping: a robust approach toward perceptually consistent mesh segmentation. In: Proc. CVPR, pp. 2145–2152 (2011)
14. Aubry, M., Schlickewei, U., Cremers, D.: The wave kernel signature: a quantum mechanical approach to shape analysis. In: Proceedings of computational methods for the innovative design of electrical devices, pp. 1626–1633 (2011)
15. Rustamov, R.M.: Laplace-Beltrami eigenfunctions for deformation invariant shape representation. In: Proceedings of symposium on geometry processing, pp. 225–233 (2007)
16. Reuter, M., Wolter, F., Peinecke, N.: Laplace-Beltrami spectra as 'Shape-DNA' of surfaces and solids. Comput. Aided Design **38**(4), 342–366 (2006)
17. Fang, Y., Sun, M., Ramani, K.: Temperature distribution descriptor for robust 3D shape retrieval. In: Proceedings of workshop on non-rigid shape analysis and deformable image alignment, CVPR (2011)
18. EL Khoury, R., Vandeborre, J.-P., Daoudi, M.: Indexed heat curves for 3D-model retrieval. In: Proceedings of ICPR (2012)

19. Rosenberg, S.: The Laplacian on a Riemannian manifold. Cambridge University Press, Cambridge (1997)
20. Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Numerical geometry of non-rigid shapes. Springer Berlin (2008)
21. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
22. Abdel-Hakim, A.E., Farag, A.: CSIFT: a SIFT descriptor with color invariant characteristics. In: Proceedings of CVPR, pp. 1978–1983 (2006)
23. Bronstein, A.M., Bronstein, M.M., Guibas, L.J., Ovsjanikov, M.: Shape google: seometric words and expressions for invariant shape retrieval. ACM Trans. Graph. **30**(1) (2011)
24. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: Proceedings of CVPR, pp. 2169–2178 (2006)
25. Shi, Y., Lai, R., Krishna, S., Dinov, I., Toga, A.W.: Anisotropic Laplace-Beltrami eigenmaps: bridging reeb graphs and skeletons. In: Proceedings of CVPR Workshops, pp. 1–7 (2008)
26. Funkhouser, T.A., Min, P., Kazhdan, M.M., Chen, J., Halderman, J.A., Dobkin, D.P., Jacobs, D.P.: A search engine for 3D models. ACM Trans. Graph. **22**(1), 83–105 (2003)
27. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE Trans. Pattern Anal. Mach. Intell. **24**(4), 509–522 (2002)
28. Kortgen, M., Park, G.-J., Novotni, M., Klein, R.: 3D shape matching with 3D shape contexts. In: The 7th central European seminar on computer graphics (2003)
29. Osada, R., Funkhouser, T.A., Chazelle, B., Dobkin, D.P.: Shape distributions. ACM Trans. Graph. **21**(4), 807–832 (2002)
30. Kazhdan, M.M., Funkhouser, T.A., Rusinkiewicz, S.: Rotation invariant spherical harmonic representation of 3D shape descriptors. In: Proceedings of symposium on geometry processing, pp. 156–165 (2003)
31. Elad, A., Kimmel, R.: On bending invariant signatures for surfaces. IEEE Trans. Pattern Anal. Mach. Intell. **25**(10), 1285–1295 (2003)
32. Ben Hamza, A., Krim, H.: Geodesic matching of triangulated surfaces. IEEE Trans. Image Process. **15**(8), 2249–2258 (2006)
33. Jain, V., Zhang, H., Kaick, O.V.: Non-rigid spectral correspondence of triangle meshes. Int. J. Shape Model. **13**(1), 101–124 (2007)
34. Coifman, R., Lafon, S.: Diffusion maps. Appl. Comput. Harmonic Anal. **21**(1), 5–30 (2006)
35. Levy, B.: Laplace-Beltrami eigenfunctions: towards an algorithm that understands geometry. In: Proceedings of IEEE international conference on shape modeling and applications, pp. 13–20 (2006)
36. Bronstein, M.M., Bronstein, A.M.: Shape recognition with spectral distances. IEEE Trans. Pattern Anal. Mach. Intell. **33**(5), 1065–1071 (2011)
37. Fouss, F., Pirotte, A., Renders, J., Saerens, M.: Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation. IEEE Trans. Knowl. Data Eng. **19**(3), 355–369 (2007)
38. Lipman, Y., Rustamov, RM., Funkhouser, T.A.: Biharmonic distance. ACM Trans. Graph. **29**(3) (2010)
39. Zhang, H., Kaick, O.V., Dyer, R.: Spectral mesh processing. Comput. Graph. Forum **29**(6), 1865–1894 (2010)
40. Sivic, J., Zisserman, A.: Video Google: a text retrieval approach to object matching in videos. In: Proceedings of ICCV, pp. 1470–1477 (2003)
41. Gemert, J.V., Snoek, C.G.M., Veenman, C.J., Smeulders, A.W.M., Geusebroek, J.: Comparing compact codebooks for visual categorization. Comput. Vis. Image Underst. **114**(4), 450–462 (2010)
42. Boiman, O., Shechtman, E., Irani, M.: In defense of nearest-neighbor based image classification. In: Proceedings of CVPR, pp. 1–8 (2008)
43. Philbin, J., Chum, O., Isard, M., Sivic, J., Zisserman, A.: Lost in quantization: improving particular object retrieval in large scale image databases. In: Proceedings of CVPR, pp. 1–8 (2008)
44. Gemert, J.V., Veenman, C.J., Smeulders, A.W.M., Geusebroek, J.: Visual word ambiguity. IEEE Trans. Pattern Anal. Mach. Intell. **32**(7), 1271–1283 (2010)
45. Yang, J., Yu, K., Gong, Y., Huang, T.S.: Linear spatial pyramid matching using sparse coding for image classification. In: Proceedings of CVPR, pp. 1794–1801 (2009)
46. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T.S., Gong, Y.: Locality-constrained linear coding for image classification. In: Proceedings of CVPR, pp. 3360–3367 (2010)
47. Perronnin, F., Dance, C.R.: Fisher kernels on visual vocabularies for image categorization. In: Proceedings of CVPR, pp. 1–8 (2007)
48. Jégou, H., Douze, M., Schmid, C., Perez, P.: Aggregating local descriptors into a compact image representation. In: Proceedings of CVPR, pp. 3304–3311 (2010)
49. Jégou, H., Perronnin, F., Douze, M., Sánchez, J., Pérez, P., Schmid, C.: Aggregating local images descriptors into compact codes. IEEE Trans. Pattern Anal. Mach. Intell. **34**(9), 1704–1716 (2012)
50. Picard, D., Gosselin, P.: Improving image similarity with vectors of locally aggregated tensors. In: Proceedings of ICIP, pp. 669–672 (2011)
51. Bronstein, A.M., Bronstein, M.M.: Spatially-sensitive affine-invariant image descriptors. In: Proceedings of ECCV, pp. 197–208 (2010)
52. Savarese, S., Winn, J.M., Criminisi, A.: Discriminative object class models of appearance and shape by correlatons. In: Proceedings of CVPR, pp. 2033–2040 (2006)
53. Ling, H., Soatto, S.: Proximity distribution kernels for geometric context in category recognition. In: Proceedings of ICCV, pp. 1–8 (2007)
54. Behmo, R., Paragios, N., Prinet, V.: Graph commute times for image representation. In: Proceedings of CVPR, pp. 1–8 (2008)
55. Liu, D., Hua, G., Viola, P.A., Chen, T.: Integrated feature selection and higher-order spatial feature extraction for object categorization. In: Proceedings of CVPR, pp. 1–8 (2008)
56. Grauman, K., Darrell, T.: The pyramid match kernel: Discriminative classification with sets of image features. In: Proceedings of ICCV, pp. 1458–1465 (2005)
57. Cao, Y., Wang, C., Li, Z., Zhang, L., Zhang, L.: Spatial-bag-of-features. In: Proceedings of CVPR, pp. 3352–3359 (2010)
58. Yang, Y., Newsam, S.: Spatial pyramid co-occurrence for image classification. In: Proceedings of ICCV, pp. 1465–1472 (2011)
59. Zhang, Y., Jia, Z., Chen, T.: Image retrieval with geometry-preserving visual phrases. In: Proceedings of CVPR, pp. 809–816 (2011)
60. Jia, Y., Huang, C., Darrell, T.: Beyond spatial pyramids: receptive field learning for pooled image features. In: Proceedings of CVPR, pp. 3370–3377 (2012)
61. Krapac, J., Verbeek, J.J., Jurie, F.: Modeling spatial layout with fisher vectors for image categorization. In: Proceedings of ICCV, pp. 1487–1494 (2011)
62. Meyer, M., Desbrun, M., Schröder, P., Barr, A.: Discrete differential-geometry operators for triangulated 2-manifolds. In: Visualization and Mathematics III, Springer, Berlin, pp. 35–57 (2003)
63. Wardetzky, M., Mathur, S., Kälberer, F., Grinspun, E.: Discrete Laplace operators: no free lunch. In: Proceedings of Eurographics symposium on geometry processing, pp. 33–37 (2008)
64. Belkin, M., Sun, J., Wang, Y.: Discrete laplace operator on meshed surfaces. In: Proceedings of SCG, pp. 278–287 (2008)
65. Hildebrandt, K., Polthier, K.: On approximation of the Laplace–Beltrami operator and the Willmore energy of surfaces. Comput. Gr. Forum **30**(5), 1513–1520 (2011)

66. Davies, E.B., Safarov, Y. (eds.): Spectral theory and geometry. Cambridge University Press, Cambridge (1999)
67. Vaxman, A., Ben-Chen, M.,, Gotsman, C.: A multi-resolution approach to heat kernels on discrete surfaces. ACM Trans. Graph. **29**(4) (2010)
68. Uhlenbeck, K.: Generic properties of eigenfunctions. Am. J. Math. **98**(4), 1059–1078 (1976)
69. Jarvelin, K., Kekalainen, J.: IR evaluation methods for retrieving highly relevant documents. In: Proceedings of SIGIR, pp. 41–48 (2000)
70. Lian, Z., Godil, A., Fabry, T., Furuya, T., Hermans, J., Ohbuchi, R., Shu, C., Smeets, D., Suetens, P., Vandermeulen, D., Wuhrer, S.: SHREC'10 track: non-rigid 3D shape retrieval. In: Proceedings of Eurographics/ACM SIGGRAPH Sympo. 3D Object Retrieval, pp. 101–108 (2010)
71. Lian, Z., Godil, A., Bustos, B., Daoudi, M., Hermans, J., Kawamura, S., Kurita, Y., Lavoue, G., Nguyen, H.V., Ohbuchi, R., Ohkita, Y., Ohishi, Y., Reuter, F.P.M., Sipiran, I., Smeets, D., Suetens, P., Tabia, H., Vandermeulen, D.: SHREC '11 track: Shape retrieval on non-rigid 3D watertight meshes. In: Proceedings of Eurographics/ACM SIGGRAPH symposium on 3D object retrieval, pp. 79–88 (2011)